

DOCUMENT RESUME

ED 316 427

SE 051 258

AUTHOR Truxal, John G.
TITLE Digital Signals.
INSTITUTION Council of Independent Colleges, Washington, D.C.
SPONS AGENCY EXXON Education Foundation, New York, N.Y.
PUB DATE 87
NOTE 72p.
AVAILABLE FROM Council of Independent Colleges Publications
Department, Technology Project, One Dupont Circle,
Suite 320, Washington, DC 20036.
PUB TYPE Reference Materials - General (130) -- Guides -
Classroom Use - Materials (For Learner) (051)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Science; Communications; Community Colleges;
Higher Education; *Information Transfer; *Nonmajors;
*Science and Society; Science Education; Secondary
Education; Secondary School Science; *Technological
Advancement; *Technology

ABSTRACT

Technological advances during the past few decades have revolutionized many complex systems that influence human activity. As the rate of technological progress accelerates, these systems will become more complex, and new ones will evolve. Citizens in a technological society need to be able to make intelligent choices about how technology will affect their lives. This requires a better understanding of the technology itself; its potential and its limitations. This is a module for a course designed to introduce students, particularly those not majoring in the sciences or mathematics, to technological concepts. This module explains the conversion of signals (audio, video, etc.) to digital form (a sequence of zeros and ones). The two primary examples are bar codes, which play a growing role in both service and manufacturing industries, and the compact disc (CD, or digital audio). This unit requires no specific background and is appropriate for an introductory course in quantitative methods or technology studies.
(CW)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

Digital Signals

by John G. Truxal

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official DERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

S.G. Pelletier

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

This module explains the conversion of signals (audio, video, etc.) to digital form (a sequence of zeros and ones). The two primary examples are bar codes, which play a growing role in both service and manufacturing industries, and the compact disc (CD, or digital audio). The unit requires no specific background and is appropriate for an introductory course in quantitative methods or technology studies.

*Designed to further student understanding of technological issues,
this module is one in a series developed in a project on
"Technology and the Liberal Arts," which is sponsored by CIC—
the Council of Independent Colleges, Washington, D.C.*

PUBLISHED BY: CIC, THE COUNCIL OF INDEPENDENT COLLEGES



Technological advances in just the past few decades have revolutionized many of the complex systems that guide our lives—the way we travel, for example, and how we communicate with one another, heal our sick, even the way we grow our food. As the rate of technological progress accelerates, these systems will become more complex, and new ones will evolve. Technology will continue to be a preeminent force in our lives.

As citizens in a technological society, therefore, we need to be able to make intelligent choices about how technology will affect our future lives. This requires that we have a better understanding of technology itself. We need to understand its potential and limitations. Further, we need to understand the very *process* of technology, and have a sense of its methodology and rules.

The Council of Independent Colleges has been deeply involved in these issues since 1984, in a project entitled "Technology and the Liberal Arts." Working with faculty from small liberal arts colleges, CIC has sought to integrate the study of technology and technological issues into the liberal arts curriculum. Our goal is to improve the understanding of technology on the part of undergraduate students—the future citizens of the technological society.

This course module is one in a series developed in the CIC project under a grant from the Exxon Education Foundation. Each module is designed to introduce students, particularly those not majoring in the sciences, to a technological issue or concept. In the process of completing the exercises in the modules, we also hope students will develop an idea of the methodology of technological inquiry. Our goal in publishing the modules is to make teaching materials available for use in courses across the liberal arts disciplines. We are deeply indebted to the Exxon Education Foundation and its executive director, Arnold R. Shore, for supporting this endeavor.

I want to also acknowledge the many college faculty who have participated in CIC's technology project. To each, I offer my sincere thanks for pioneering the important effort to bring technology into the mainstream of the curriculum.

Allen P. Splete
President
The Council of Independent Colleges
October 1987

*In addition to the Exxon Education Foundation,
the following sources have provided funding for the
CIC project on Technology and the Liberal Arts:
Monsanto, Pew Memorial Trust, AT&T, Ford Foundation,
Consortium for the Advancement of Private Higher Education, Pfizer, and Xerox.*

Digital Signals

by John G. Truxal

Distinguished Teaching Professor
Department of Technology and Society
College of Engineering and Applied Sciences
State University of New York at Stony Brook
Stony Brook, New York

INTRODUCTION

In 490 B.C. the Athenians won the Battle of Marathon. Pheidippides ran the 26 miles back to Athens, gasped "Rejoice, we conquer," and died.

One hundred years ago John D. Rockefeller in New York City sent messages by telegraph to his refinery in Cleveland and to his oil wells in western Pennsylvania. The technology made possible daily business transactions among distant locations.

In 1986 a space vehicle traveled by the planet Uranus. From a distance of almost two billion miles, we received television pictures showing details of the planet and its satellites to provide us with new information about the universe.

The history of our civilization is closely related to our ability to communicate at a distance (telecommunication). Particularly in the last 140 years, our telecommuni-

Copyright © 1987 Council of Independent Colleges

cation systems have allowed us to become "one world." We are able to watch live television of a satellite launch in the Soviet Union, speak to friends in India, and transfer funds to banks in Australia.

Rate at which information can be sent

The change in telecommunication capability over the years has been primarily due to the *rate* at which information can be sent. In 1866, telegraph messages could be sent to England over the trans-Atlantic telegraph cable. But a telegraph signal is very slow in conveying information. With the advent of the telephone that allowed speech transmission, information was conveyed at a much higher rate. The telephone trans-Atlantic cable made it possible to send several phone conversations simultaneously. The new fiber optic cable in 1988 will allow 40,000 conversations on the single system.

Thus, telecommunications capability can be measured by how many *simultaneous* phone conversations can be sent along the same cable, pair of wires, microwave relay, or satellite relay. Figure 1 shows this plot over the last 150 years. In the 1840s, one telegraph signal carried the information of about 1/1000 of a phone conversation. A hundred years later we could send 1000 conversations, which happens to be equivalent to the information of one television picture.

Growth to 1970s

Figure 1 shows that the growth of telecommunication capacity is a straight line up to the 1970s. Of course, this straight line is somewhat deceptive since each vertical unit represents an increase by a factor of 100. About every 34 years, the capacity was multiplied by 100. (We call this exponential growth, as explained further in Problem 1.) In one person's lifetime (68 years), there are two 34-year periods, so the capacity would be multiplied by 100×100 or 10,000. We have seldom seen such phenomenal growth rates in any human activity. What caused or allowed such growth?

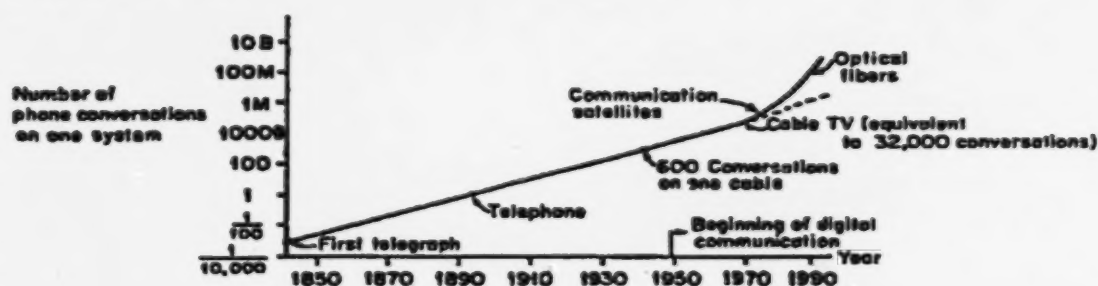


Figure 1 — Growth of telecommunication capacity over the last 140 years.

There is no simple answer to this question. Growth of telecommunication capacity went hand-in-hand with the emergence of large corporations — which caused which is irrelevant; neither could have happened without the other. For example, as telecommunication capacity grew, businesses found that they could operate at widely scattered locations and still communicate to maintain a well-organized company. Corporate headquarters were set up in New York City to be close to banks and financial centers. Manufacturing plants could be spread throughout the country to take advantage of lower labor costs, convenient supply of raw materials and electrical energy, and proximity to the buyers of finished products.

Thus, telecommunication growth is closely tied to this country's economic development. This spreading out of parts of businesses to optimum locations still goes on. As a simple example, if you are on the phone in New York City and dial 411 to reach "directory assistance," you are probably connected automatically to a human operator in Rochester, New York — about as far away as possible within the state. The phone company has found it is cheaper to use operators there, where labor costs are lower; the savings are much greater than the extra costs of the long-distance phone calls.

As a more interesting example, we turn to the economic problems of New York City. The city lost a large percentage of its manufacturing industry because of high taxes, high wage scales, transportation problems, and so on. In the post-World War II years financial institutions (banks, brokerage houses, insurance companies, etc.) replaced the manufacturing industries. In recent years, however, these companies have been moving their "back-office" or paperwork operations far from the city. A company can have all the routine paperwork done in Oklahoma City, for example, where rents, taxes, and wages are lower and where there is a better supply of qualified clerical people. The cost of telecommunications is so low, the location of the "back-office" is unimportant: it costs almost nothing to transmit records of stocks bought and sold.

Will the headquarters of the financial institutions move, too, leaving in New York City only the brokers on the floor of the stock exchange (until the buying and selling is entirely electronic)? In an attempt to keep New York the financial capital of the world, the City, the Port Authority of New York and New Jersey, and Merrill Lynch are building a "teleport" on Staten Island.

Here 18 antennas will be aimed toward different communication satellites hovering in geostationary orbit over the Americas. Fiber optic cables will carry the signals from Staten Island to downtown New York (Wall Street in lower Manhattan) and to corporate headquarters throughout the city. A corporation will be able to rent as many channels (as much communication capacity) as it needs to ensure immediate communication with distant offices. A smaller company will be able to share use of a channel (at lower cost).

Such economic incentives certainly contributed to our telecommunication

growth. There were social incentives and pressures as well. The radio networks that emerged in the 1920s and the television networks of the 1940s required telecommunication capability to bring programs to broadcasting stations over the country. Today most TV commercials are filmed in New York, then transmitted via satellite for broadcast within regular programs. The public demand for radio, television, and phone service has made "necessities" of these three types of telecommunication.

In this communications case, the changes in technology have resulted from economic and social pressures. Historians often state that technological change follows scientific discovery: the scientist finds a new phenomenon, then the engineer develops the technology which uses that science. While this science→ technology sequence does occur, there are many cases in communications where the technology preceded the science. For example, we had transcontinental telephoning in the 1920s, before the scientific explanation of why it worked.

Thus, there is no *simple* explanation for the awesome growth shown in Figure 1 up to the 1960s.

Growth since 1970

Figure 1 shows that growth of telecommunications capacity during the last 17 years is even faster than exponential. What caused this acceleration?

To answer this question, we must discuss *signals*. A signal is any quantity that can be measured; when we communicate, we send a signal to the receiver in our communication system.

For example, your heart rate is perhaps 60 beats/minute when you are resting. During excitement or strenuous activity, your heart rate may rise to 160 beats/minute as the body demands higher blood pressure and more blood flow. If you were an astronaut walking on the moon, the NASA control center in Houston would be monitoring your heart rate: an instrument on your wrist would measure the rate, and a radio transmitter would send this signal (Figure 2) to Houston.

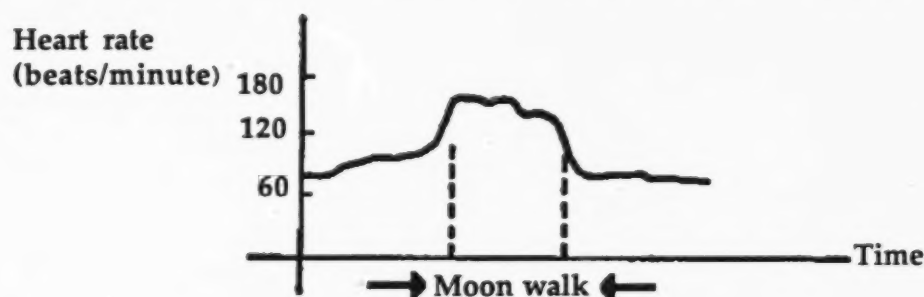


Figure 2 — Heart rate of an astronaut on the moon.

The heart rate is a signal: it carries information from the source (your wrist monitor) to the receiver or observer (the doctor in Houston). Similarly, when you are talking to a friend, the sound signal traveling from your mouth to his ear changes continually to convey information. A signal is the basis of communication — what we want to communicate.

Now let us return to our earlier question: why has our communication capability grown so rapidly in the last 17 years? To a large extent, the answer is because we have begun sending *digital signals*. In other words, instead of radioing the heart rate of Figure 2 continuously (at every instant of time), we send a long sequence of binary digits (a binary number system has only two symbols, usually 0 and 1):

010011101000010 . . .

This sequence of digits tells Houston what your heart rate is every few seconds (more than often enough). For example, the first 10 digits might mean a rate of 68, the next ten digits 72, and so forth.

Today almost all communication technologies use digital signals. In an automobile assembly plant, automated robots learn which car is moving in front of them through a digital signal from the control computer. The master controller or “brain” of the Bay Area Rapid Transit system in San Francisco sends a digital signal to each train to control its speed. Digital audio (CD’s or compact discs) and digital television promise to dominate the field of electronic entertainment.

Thus, we are living and working in a world of *digital information* — a world in which information moves in the form of digital signals. We need first to appreciate what is meant by the term *digital signals*. In this module we hope to present the basic concepts underlying this technological revolution.

The advantage of putting information into digital form is so great that we are convinced any advanced civilization trying to communicate with us would understand this form. Thus, when we sent a tablet into space in the hope it might eventually reach another civilization, we used a digital message (see Problem 2). When we point our listening antennas toward a likely spot in the galaxy where there might be an advanced civilization (the SETI program, or the Search for Extra-Terrestrial Intelligence), we assume any incoming message will be in digital form. A major goal of this module is to show why digital signals are so dominantly desirable.

1. A Tactile Communication System

In this section, we start with a simple example of *digital signals*. You and I are

together in a noisy and dark environment, and I want to send you a message. If I speak, you can not hear me because of the noise; if I move my hands to signal, you can not see in the darkness. We decide to use the sense of touch — tactile communication. We agree ahead of time on a code, and I touch you at pre-selected points to indicate the message.

Now let's turn to a simple tactile communication system. As shown in Figure 3, I stand behind you and face your back. I want to communicate to you three numbers

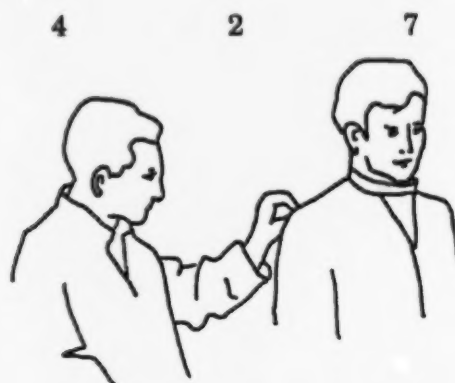


Figure 3 — A simple tactile (touch) communication system

We agree ahead of time that I will touch you on an imaginary line across your back — with the far left end representing a 0, the far right a 9, and the other digits at appropriate intermediate points (Figure 4).

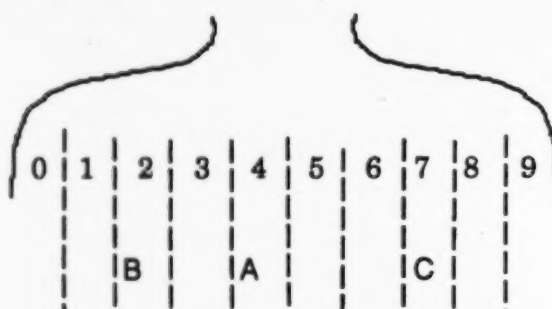


Figure 4 — The code or arrangement we agree on for our tactile communication system. To send you the information 4 2 7, I touch you, in turn, at points A, B, and C.

This is not a good communication system. The chances are that you will decide

the numbers I sent were: 4 2 6 or 4 3 8 or 3 2 9

You will seldom receive the correct message. Even if I repeat the sequence of three digits several times, the reliability is not good. The trouble is that each time I touch you, there are too many possible locations (10) from which you must decide on one. We are trying to send too much information with each touch.

Let's try a different scheme. First, we'll convert each digit (4, 2, and 7) to binary form according to the conventional binary numbers.

<u>Decimal number</u>	<u>Binary number</u>
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

You and I agree on this code ahead of time.

Now to send you the number 427, I will send the 12 binary digits.

0100 0010 0111

For a 0, I'll touch you on the left side of your back, for a 1 on the right side. Thus, I'll touch you 12 times in sequence:

LRLl LLRL LRRR

Communication is slower, but the reliability is great! All you have to decide is whether I touched you on the left or right side each time. As long as I stay well

away from the middle of your back, there is really no chance of error. We have a communication system which is slow, but *hardly ever has an error*.

If we are worried that there might be a single error occasionally, we can even send one more *check digit*. You and I agree ahead of time that after I've sent 12 binary digits, I will count the number of ones. If this number is even, I'll send a zero (to make the total number of ones stay even). If the first 12 digits have an odd number of ones, I'll send a one to leave an even number of ones:

0100 0010 0111 1

When you receive this 13-digit message, you count the number of ones. If your total is even, you know a single error has not occurred. (There might be two errors, but we assume the probability of that is negligible.) You know the message you have is correct.

Thus, we have a communication system which is in practice essentially perfect. There is almost no chance of an error; if there should be one error, you recognize this right away.

2. Comments on Our Tactile Communication System

We have devised a remarkably reliable, tactile communication system. (The word *tactile* simply means we are using touch.) Why is the system so free of errors? Primarily because the receiver (you) need only determine which of *two* signals was sent. For each binary digit (0 or 1, left or right) you don't care exactly how far to the left or right the signal is, but rather only if it is to the left or right.

If you turn slightly as I am about to touch your back, there still is no error, as long as your turning does not move my contact from the left half to the right. *Noise is any unwanted disturbance (or signal) added to the signal carrying the information.* In our system noise can be your turning slightly, the wind moving my arm or my inadvertently jerking sideways. In our digital system, a small amount of *noise* does not cause an error.

The system has a second important feature: we can easily add an extra digit or a few extra digits to check the accuracy of the communication. We call this use of extra check digits *redundancy*. After we send a block of information or message digits (12 in the example above), we stop the message for a check digit (the redundant or extra digit). This concept was first proposed during World War II by van Duuren; since then, the use of redundancy to improve accuracy has become almost universal in communication. (In Section 8, we discuss redundancy more fully.)

Finally, there is a third advantage of the binary communication system: the transmitting and receiving equipment is simple. In our human system, I need only move my hand well into the left or right half of your back; the exact location is immaterial, so I do not need precise operation of my nervous and muscular systems. In an electronic system, the devices could be quite crude but operate satisfactorily.

Thus, there are three major advantages of the binary communication system:

- (1) Very few errors
- (2) Easy to insert redundancy for detection of errors
- (3) Simple equipment

What do we pay for these advantages? A longer message. Instead of three decimal digits (427), we have to send at least twelve binary digits (010000100111). This disadvantage diminishes when we study machine systems, since electronic equipment often can send a zero or one much more rapidly than a precise decimal value. Even with this feature, however, we do need more time to send binary signals.

The advantage of digital signals are so overwhelming that essentially all communication technologies are changing to this form. As we will see in the appendix, CD's or compact discs use digital signals to record music. Digital television is about to take over home TV. Even our well-established telephone system is changing rapidly. Finally, more and more of our telecommunication capability is used by computers "talking" to computers, and computer information is always digital.

3. Morse Code

The Morse Code is an example of a signal which is basically digital. As the insert (next page) on the history of telegraphy shows, there were operating telegraph systems well before Morse's 1840 demonstration. Morse's contribution was the simplification and accuracy achieved by using a system of dots and dashes to represent each letter.

In the Morse system, an electrical current flows from the transmitter through wires to the receiver. When the current flows for only one interval of time, we are sending a dot; a dash lasts three intervals of time. Thus, the letter A is represented by

Dot	Dash
-----	------

with one interval of time between the two symbols. In terms of binary numbers, the signal is

1 0 1 1 1

Dot	Dash
-----	------

Thus, the system has the advantages inherent in digital signals: in each time interval, the receiver need only determine whether there is a current or no current.

The Morse code represents a major technological advance for another reason: the code is quite "efficient." That is, the more frequently the letter appears in written English, the faster it can be sent. For example, E is by far the most common letter (Table 1), so Morse chose a single dot to represent an E. The second most

HISTORY OF THE TELEGRAPH

- 1790 Claude Chappe, France, developed a semaphore communication system which spread through Europe, then by 1838 as far as Moscow. This was advanced technology before electricity.
- 1809 Soemmering demonstrated a crude electric telegraph—at the receiver, electricity produced gas at one of 35 electrodes.
- 1820 Ampere, France, demonstrated a telegraph system with 26 wires, one for each letter, with the original suggestion by Laplace.*
- 1837 Cooke and Wheatstone, England, patented a telegraph system quickly adopted by European governments.** The electric current transmitted determined how far a needle at the receiver would rotate and hence indicated a specific letter.
- 1840 Alfred Vail joined Samuel F. B. Morse and demonstrated a telegraph with the major advantage of a binary signal—dots and dashes of the Morse code.
- 1843 At the last session, Congress approved \$30,000 for a demonstration of the telegraph.
- 1844 Morse sent his famous message, "What hath God wrought," from Washington to Baltimore, where Vail sent it back.
- 1844 Summer—news that Polk had won the Democratic Presidential nomination in Baltimore reached Washington by telegraph.
- 1845 Public enthusiasm resulted when a murderer escaped police in Slough, England and boarded a train. A description telegraphed ahead allowed police to arrest him.
- 1851 There were more than 50 telegraph companies in the U. S.
- 1856 Western Union Telegraph Company emerged.
- 1861 October 24: The first transcontinental message assured President Lincoln that California would stay in the Union. (The Pony Express died shortly thereafter after operating only 19 months).
- 1866 The first trans-Atlantic telegraph line began operation (by 1884, six more were in place).

The speed with which telegraph covered the U. S. was awesome: only 17 years after the Morse demonstration, transcontinental messages were sent.**

*Like so many mathematical leaders, Laplace was a child genius. A professor at 18, six years later he proved that the observed motion of the planets did not require occasional divine intervention, as Newton had thought. Later, Napoleon appointed him Minister of the Interior, but fired him in six weeks for bringing the "spirit of infinitesimals into administration."

**In 1972, Alvin Toffler's book *Future Shock* argued that the increasing rate of technological change is reaching a level where human beings and social institutions cannot adapt. Even in the White House, the question was raised on the seriousness of the threat. Yet the rate at which the telegraph changed telecommunication is probably unmatched today.

common letter is T, which is represented by a single dash (requiring three times as long to send as a dot). Thus, a minimum amount of time is required to send a long message—the code is efficient.

E	132	R	68	L	34	G	20	V	9
T	104	I	63	F	29	Y	20	K	4
A	82	S	61	C	27	P	20	X	1
O	80	H	53	M	25	W	19	J	1
N	71	D	38	U	24	B	14	Q	1
								Z	1

Table 1. Frequency of letters in written English. In each 1000 letters, E appears 132 times, T 104 times, and so on.

E .	S . . .	U . .	B	P . . .
T -	R . .	L . . .	V	X
I . .	H	F	K . . .	Z
A . -	D . .	G . . .	O	Y
N - .	M . .	W . . .	C	J
				Q

Table 2. Modern Morse Code, with letters ordered according to time for transmission. Comparison with Table 1 shows that the only common letter with an inappropriately long sequence is O. (Actually Morse's original code was slightly different: O, for example, was represented by a dot - double space - dot.) The similarity between the letter sequences in Tables 1 and 2 indicates the remarkable efficiency of the Morse code. Indeed, the best possible code would save only 6 percent of the time required to send messages.

Actually, today's telegraph equipment almost never uses the Morse code. With machines to transmit and receive, the number of letters sent per second is much less important than the accuracy; consequently, we go to a digital message (a set of ones and zeros) with redundancy to allow automatic determination of an error. Each letter is represented by seven binary digits — for example,

0100011

or

1000110

With seven digits available, there are 2^7 ($2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$) or 128 different possibilities, far more than we need. To allow automatic recognition of any single error, we use only the seven-digit binary numbers *with three ones*. Now how many different, allowable numbers are there? We might simply list these and find there are 35 — plenty to cover all the letters in our alphabet (Table 3).

A 0011010	F 0010011	K 0001011	P 1001010	U 0110010
B 0011001	G 1100001	L 1100010	Q 0001101	V 1001001
C 1001100	H 1010010	M 1010001	R 1100100	W 0100101
D 0011100	I 1110000	N 1010100	S 0101010	X 0010110
E 0111000	J 0100011	O 1000110	T 1000101	Y 0010101
				Z 0110001

Table 3 — The van Duuren code, which uses seven binary digits for each letter, always with three ones. If the received signal has one error in each packet of seven digits, the error is immediately recognized (there will be either two or four ones). H.C.A. van Duuren invented the code during World War II; by 1956 it was an international standard under the title ARQ for Automatic Repeat Request — when the receiver sees other than three ones in a packet of seven digits, it automatically requests a repeat of the packet. The system is not common in satellite communications since a relay from a geosynchronous satellite involves a delay of at least 1/4 second each way. The ARQ would arrive at the sender more than a half second after the original signal.

In this section, we have tried to introduce the idea of digital signals with a very simple example: telegraph messages. The telegraph really ushered in modern communications. Messages travelled over wires at nearly the speed of light. Never again would we have a major battle fought after the war was over (the Battle of New Orleans in the War of 1812, between the signing of the peace treaty in London and the arrival by ship of the news in this country).

Braille Code

A	1	K	13	U	136
B	12	L	123	V	1236
C	14	M	134	X	1346
D	145	N	1345	Y	13456
E	15	O	135	Z	1356
F	124	P	1234		
G	1245	Q	12345	6 added to second column	
H	125	R	1235	W	2456
I	24	S	234	(French alphabet in 1825 did not include W)	
J	245	T	2345		
		3 added to first column			

Louis Braille, by the age of 15 (1825), had developed his alphabet and system of printing for the blind. He learned how Napoleon's soldiers passed messages at night by embossing on cardboard (another example of the fall-out from military work).

In Braille printing, we emboss or raise one or more of six dots for



each letter. If these dots are numbered by location

1	4
2	5
3	6

the table above shows the code. In other words, each symbol is a set of six binary digits.

There are 63 possible symbols; we use only 26 above. The others represent punctuation marks,

common letter combinations (2346 is "the"), or a sign that the next symbol is from a special alphabet (to allow other common combinations). This complexity of the code means that 1-2% of the blind population can read Braille at a rate of 100 words/minute or more (the rate which must be reached if reading of books and magazines is to be pleasurable).



4. Sampling

When we listen to music or to someone talking, what is the sound? How does the sound reach our ears?

When there is no sound, the air pressure around us is about 14.7 pounds per square inch — the “atmospheric pressure” resulting from the atmosphere over our heads. We live in this environment of air pressure which varies slightly from day to day.

A sound signal from a loudspeaker or a person talking causes this air pressure to change very, very slightly (air is compressed then expanded). Figure 5 shows the way the air pressure midway between the loudspeaker and our ears might change with sound present. Up to time t_1 , there is sound, and the pressure varies above and below its normal value. From t_1 until t_2 there is no sound, and the pressure stays at its normal value. At t_2 , the sound starts again.

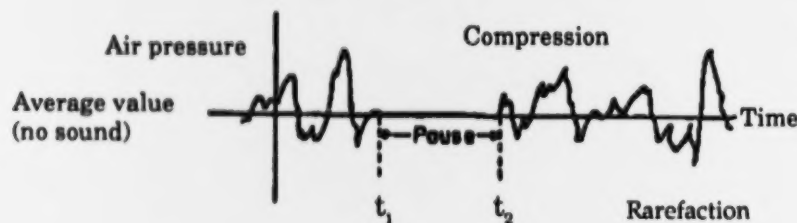


Figure 5 — A possible sound signal.

This pressure or sound signal travels from the loudspeaker to our ears, where the human hearing system involves the brain's interpretation of the sound signal. For our purposes here, however, only the sound signal is of interest — the way air pressure changes with time.

The signal shown in Figure 5 is an *analog signal*. That is, the signal exists at every instant of time and changes smoothly from one value to the next. Many signals we encounter in real life have this analog characteristic: e.g., the speed we are traveling in a car, or the outside temperature or wind velocity.

We want to describe this signal in digital form — by a sequence of zeros and ones. How can we possibly represent a sound signal like that shown in Figure 5 by a sequence of numbers?

Fortunately, we do not need to describe the signal at every instant of time. Any real-world signal never changes instantaneously from one value to another, very different value. Changes are always smooth and are limited in how fast they occur.

Thus, if we look at a small piece of the sound signal in Figure 5, we find the gradual changes shown in Figure 6—even if we pick a time interval when the signal is changing most rapidly.



Figure 6 — Small segment of the sound signal.

When any signal is limited in how fast it can change, we can describe that signal by *sampling*—looking at the signal frequently for a very short period of time. Thus, the sound signal of Figure 6 is entirely equivalent to the sequence of samples shown in Figure 7. From the actual sound signal, we can find these samples. If the samples are given, we can reconstruct the original, smooth sound signal.

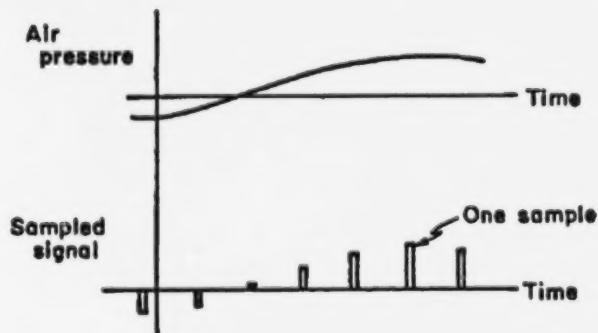


Figure 7 — Samples of the sound signal. The sample height is the size of the air-pressure signal at the instant of sampling.

Actually the idea of sampling is nothing especially novel in human experience. We sample frequently in the course of our normal activity. As I sit here writing, we are in January on Long Island and there is a threat of snow tonight. Every once in a while I glance out the window to see whether the snow has started: my decision on when to leave the office will depend on when the snow starts. I am sampling the weather-condition signal by occasionally glancing out the window.

Driving home, my eyes will frequently focus on the road in front of my car. I also sample the signal from behind through the rear-view and side mirrors, I look right and left to observe cars approaching on side streets, and so on. Indeed, engineers have used a TV camera directed toward the driver's eyes to determine that a normal driver is looking at the road ahead less than 25 percent of the time under average conditions.

We know that, if we sample frequently enough, we lose no information because of the sampling process. I do not really care exactly when it starts snowing—the precise moment the first flake hits the ground. I would like to leave for home within a half hour of the start of the snow. Consequently, if I look out the window every 15 minutes, I certainly will have all the information I can use. There is absolutely no reason to stare unblinkingly out the window.

How frequently we must sample depends on the rate at which the signal may change *significantly*. I feel ill in mid-morning and take my temperature; if it's normal, I see no reason to take it again five minutes or even two hours later. On the other hand, when driving my car, I may have to hold my eyes riveted on the road ahead if I'm following a car driven by an alcoholic, or if there are major potholes or heavy fog. In these cases, the signal (or situation) may change rapidly — I must sample frequently.

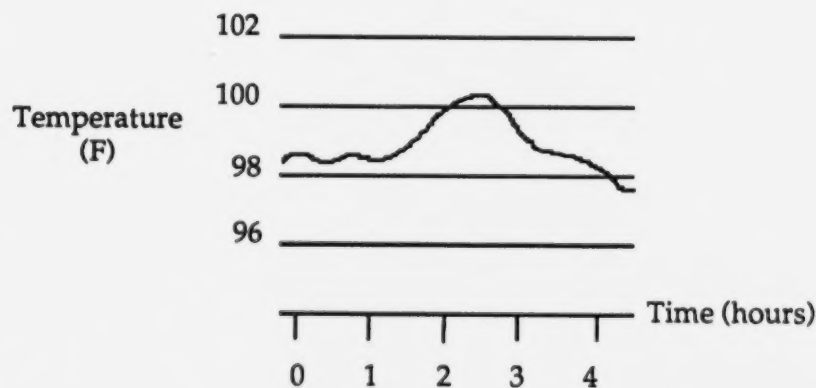


Figure 8 — Body temperature of a patient.

How often we need to sample can be discussed in intuitive terms. Figure 8 shows an analog signal: the long-term record of the body temperature of an individual from the time we start measuring (time 0) for the following four hours. Inspection of this curve reveals that the most rapid change occurs in the first 45 minutes during the second hour.

Even in this rapidly changing portion, however, nothing significant happens in less than 15 minutes. Thus, a measurement every 15 minutes provides enough data to keep the physician fully informed of the patient's condition. Instead of the curve of Figure 8, representing the temperature at *every* instant of time, we might sample and obtain the data:

<u>Time (minutes)</u>	<u>Temperature (°F)</u>
0	98.5
15	98.5
30	98.6
45	98.6
60	98.7
75	99.2
90	99.7
105	100.3
120	100.1
135	99.8
150	99.5
165	99.4
180	98.9
195	98.7
210	98.5
225	97.9
240	97.7

Table 4 — Data representing the body-temperature signal of Figure 9.

Either this table or the corresponding graph of Figure 9 is an adequate description of the signal: the way the body temperature changes as time passes.

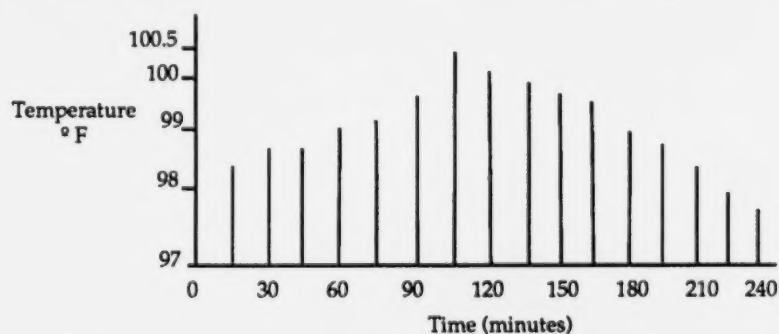


Figure 9 — Sequence of samples representing the body-temperature signal of Figure 8. We sample every 15 minutes.

Thus, we have to sample often enough to be sure that the samples show any important change in the signal. The more frequently we sample, the more complicated the sampling will be, so communication equipment usually samples at close to the minimum rate necessary.

Intuitively we can see that we have to sample often enough so that we do not miss any changes in the signal. In Figure 10, we do not sample frequently enough—our four samples miss entirely the rapid change between samples 2 and 3. Thus, to find how often we need to sample, we might look at a lengthy graph of the signal, find the interval when changes are most rapid, and choose the sampling rate accordingly.

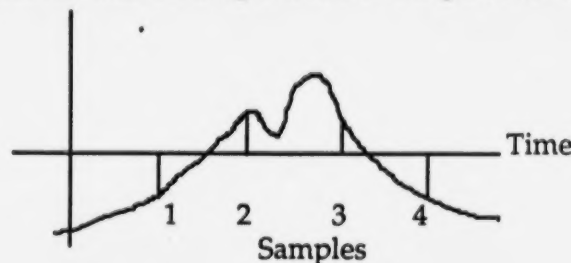


Figure 10 — Sampling too seldom

There is a mathematics theorem which tells the engineer the smallest allowable sampling rate (the Nyquist sampling theorem developed in 1933 by Harry Nyquist of Bell Laboratories). We can not prove or really explain the theorem in this module, since it requires mathematical background which we don't want to develop. (See Appendix A for a brief discussion.)

Regular Sampling

In the example of Figure 9, the samples appear regularly, every 15 minutes. When people sample in daily activities, we tend to adjust the sampling rate according to the rate the signal is changing. In driving my car, I glance in the rear-view mirror occasionally. If there is a police car or emergency vehicle behind me, I increase the sampling rate enormously since I am now interested in rapid changes in the situation. Similarly, if the patient's body temperature passes 104°F , the physician wants to know immediately if there is a further rise.

Technological systems usually work with a constant sampling rate. We sample at regular intervals — e.g., every $1/8000$ of a second for a phone conversation. Even if the speaker pauses for a half second between words, the sampling rate remains the same. The equipment is enormously simpler if we sample at regular times, even though nothing may be happening during many of the samples.

Final comment on sampling

Sampling is the first step in converting the signal to digital form. Sampling gives a signal where the values are known at particular, regular instants of time.

We have two more steps before the signal appears as a sequence of zeros and ones (binary digits).

5. Quantizing

We are in the middle of converting a signal to *digital* form—that is, representing all the information in the signal by a sequence of binary digits. Figure 11 shows the process; we have moved from A to B at this point.

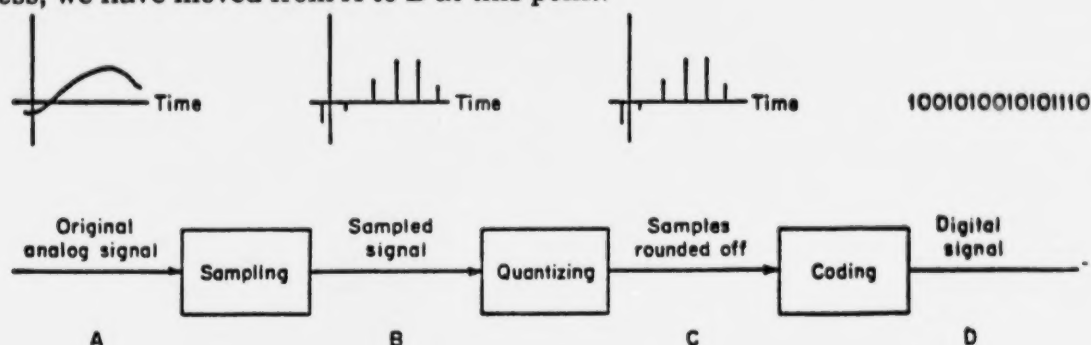


Figure 11—Three processes in converting a signal to digital form.

The second step is *quantizing*, which means rounding off the sizes of the samples. For example, if human body temperature is the signal, we might have a sequence of samples at B in Figure 11.

97.313
98.219
100.172
101.861
101.873
100.444

Clearly, these numbers are much more precise than any doctor would want; we should round off the numbers.

We need to decide how precise to be. Suppose we find that a half degree is sufficient for medical purposes. Then we change the numbers to:

97.5

98.0

100.0

102.0

102.0

100.5

This rounding-off carries the name "quantizing." (A quantum is a minimum chunk; here any change must be a minimum amount of 0.5°). Thus, at point C in Figure 11, the sampled/quantized signal looks graphically much the same as the sampled signal at B, before quantizing, but the sizes are all rounded off.

Obviously we do introduce some error in this quantizing process. 97.5° is not the same as 97.313° . But even our original set of sample values is quantized, since every thermometer has a maximum number of decimal places that it reads. While we do discard information in rounding off, no information of significance is lost.

In a sense, this quantizing throws away irrelevant, confusing detail. We do this all the time to allow intelligent thinking. When we say that in the last 20 years the population of New York City has dropped from eight million to seven million, we realize that these are only rough numbers — but they are precise enough to describe the important aspect.

6. Coding

In Figure 11, showing our three-step process for changing to a digital signal, we are now at point C—ready to do the coding or conversion to binary numbers. If we continue our example of the body temperature, we have the measurements

97.5 98 100 102 102 100.5

(Recall that we have rounded off to 0.5° .)

The first question to ask is: With the rounding off we have done, how many different signal values can there be? In this example, we are measuring human body temperature, which in almost every case is between 95° and 107° . There are then only 25 different temperatures we may encounter

95	95.5	96	96.5	97	97.5	98	98.5	
99	99.5	100	100.5	101	101.5	102	102.5	
103	103.5	104	104.5	105	105.5	106	106.5	107

We are planning to represent each of these temperatures by a different binary number (e.g., 01001). How many digits (zeros or ones) do we need in each binary number to allow at least 25 different possibilities?

If we use just one binary digit*, we find the results of Table 5.

<u>Number of binary digits</u>	<u>Total number of sample levels</u>
1	2
2	4
3	8
4	16
5	32
6	64
7	128
10	1,024 (about one thousand)
20	1,048,576 (about one million)

Table 5 — The binary system is not efficient in terms of the number of digits required. To represent decimal numbers up to 999, we need 10 binary digits.

Hence, to represent 25 different temperatures, we need five binary digits or bits. (Four digits would give only 16 possibilities). Now we can choose our *code*, or the rules by which each binary number represents a different temperature (Table 6).

95 — 00000	97.5—00101	100— 01010	102.5—01111	105 — 10100
95.5—00001	98— 00110	100.5—01011	103 — 10000	105.5—10101
96 — 00010	98.5—00111	101 — 01100	103.5—10001	106 — 10110
96.5—00011	99 — 01000	101.5—01101	104 — 10010	106.5—10111
97 — 00100	99.5—01001	102 — 01110	104.5—10011	107 — 11000

Table 6 — One possible coding scheme. Each temperature is represented by a unique, five-digit binary number.

*Instead of always writing the term "binary digits," we can adopt the vocabulary of the engineer: the phrase "binary digit" is abbreviated "bit" (the bi from binary, the t from the end of digit).

The sequence of temperatures we measured corresponds to the digital signal:

00101	00110	01010	01110	01110	01011
97.5	98	100	102	102	100.5

We have completed the conversion from the original analog signal to a digital signal.

The three steps (sampling, quantizing, and coding) might be called “digitization” — conversion to digital form. Engineers have a special term: *pulse code modulation*, abbreviated PCM. Modulation means a change in form, here from an analog to a digital signal. The adjective “pulse” means our final signal uses a sequence of pulses (e.g., a positive pulse for a one, no pulse for a zero). The adjective “code” indicates that we code the signal.

There is one final comment: The code we chose above (Table 6) is very simple. We just listed the binary numbers in order in the usual way of counting; thus the successive binary numbers are the representations of 0, 1, 2, 3, . . . Then we let these numbers correspond to successive values of temperature, starting with the smallest. This decision was arbitrary. We could have arranged the binary numbers in any order. For example, we might choose the sequence:

00000	00001	00011	00010	00110	00111	00101	00100
01100	01101	01111	01110	01010	01011	01001	01000
11000	11001	11011	11010	11110	11111	11101	11100
							10100

Known as the Gray code, this arrangement is occasionally advantageous because each entry is only one digit different from the preceding or following entry. Since the temperature is most likely to stay the same or to change at most 0.5° from one sample to the next, this code may result in simple equipment.

We only want to emphasize here, however, that the code chosen is arbitrary. At the end of this module, we show another code used for speed control in the BART subway system as a second example of this flexibility.

7. Breaking the USPS Code

With the U.S. Postal Service moving toward automation in letter sorting and routing, we find arriving letters now frequently carry the printed bar code



Presumably these 32 symbols represent 11794 (the Stony Brook zip code). As a problem, we want to break the code. Initially we work only from this one example; as work progresses, we will use other examples.

One possible reasoning sequence follows:

(1) The bars are of two heights, so probably the code is binary. If we let a short bar represent a *zero*, a tall bar a *one*, 11794 is

10001100011100011010001001100101

(2) Perhaps each decimal digit is represented by its binary equivalent:

0 — 0000	5 — 0101
1 — 0001	6 — 0110
2 — 0010	7 — 0111
3 — 0011	8 — 1000
4 — 0100	9 — 1001

If so, the two 1's starting 11794 should give the binary sequence

00010001

but this does not appear near the beginning of the envelope code.

(3) There are 32 symbols to represent the five decimal digits. There is no obvious reason to make the code complex — cheating the code does not seem to be a rewarding pastime. Consequently, each decimal digit is probably represented by the same number of bits (binary digits). But 32 is not divisible by five.

Therefore, we might guess the initial and final 1s are just "guard rails" — that the 30 interior digits represent 11794. If so each six bits represent a decimal digit:

1	000110	001110	001101	000100	110010	1
Start	1	1	7	9	4	End

But this is not attractive. If the code is simple, the bits for a 1 probably are the same, regardless of where the 1 appears in the zip code.

(4) Perhaps there are six bits to start and end (again we assume symmetry), and four bits for each decimal digit

100011	0001	1100	0110	1000	1001	100101
Start	1	1	7	9	4	End

Again the two 1's have different bit representations.

(5) Maybe a check digit is used, so there are six decimal digits to be represented. Again, if there is one bit to start and end, each decimal digit is represented by five bits:

1	00011	00011	10001	10100	01001	10010	1
Start	1	1	7	9	4	Check	End

Finally, this looks good: at least both 1's have the same five bits.

(6) If we assume we are on the right track, we know four representations

1	00011
7	10001
9	10100
4	01001
Check for 11794	10010

All five of these have exactly two 1's. Perhaps each decimal digit is represented by a five-bit sequence with two 1's. How many possibilities are there?

11000	01100	00110	00011
10100	01010	00101	
10010	01001		
10001			

Exactly the ten that we need to represent 1, 2, 3, 4, 5, 6, 7, 8, 9, 0.

(7) At this point we need an additional example. My home zip code is 11746, which is represented by



or

1	<u>00011</u>	<u>00011</u>	<u>10001</u>	<u>01001</u>	<u>01100</u>	00011	1
Start	1	1	7	4	6	Check	End

This tells us that 6 is 01100.

In addition, it indicates that the check digit for 11746 is a 1. Probably the rule for finding the check digit is simple. Maybe I just add the five zip-code digits:

$$1 + 1 + 7 + 4 + 6 = 19$$

An error in the last digit of the sum would indicate a single error in one of the digits. But my last digit is a 9, the check digit is 1. Perhaps the check digit is tens-complement of the last digit of the sum: that is, whatever I must add to the sum to make the last digit zero.

If this is the correct rule, the check digit for 11794 would be 8: the sum is 22, 100 - 22 is 78, and the last digit is 8. If this is correct,

8 is 10010

(8) We now know

1	00011
4	01001
6	01100
7	10001
8	10010
9	10100

but we still do not know 0, 2, 3, 5. If we look at the above listing, we find that each number is larger than the one above in regular binary numbers. Suppose we list all

five-bit numbers with two ones (from paragraph 6) in order of increasing size

00011 - 1	01100 - 6
00101	10001 - 7
00110	10010 - 8
01001 - 4	10100 - 9
01010	11100

Probably the others are in the order listed (recall we assumed the code is simple):

00101 - 2
00110 - 3
01010 - 5
11000 - 0

and perhaps we have broken the code.

A mailing from *American Scientist* in New Haven arrives with its nine-digit zip code 06511-9988



We are delighted to see this seems to confirm our analysis.

This example shows how bar codes can be helpful in automation of both manufacturing and service industries. In the postal system, the post office must add the bar code. When you mail a letter, it goes to the local distribution center. There the letters travel one at a time on a conveyor belt past a clerk who reads the zip code and types in this code so that a machine can print it on the bottom right of the envelope. Thereafter, automatic equipment can sort and route the letter to the recipient's local post office. If the nine-digit code is used, equipment can sort mail for the local carrier.

When the zip code was first introduced, the hope was that equipment would be able to read the printed or typed zip code automatically. This has proven to be extremely difficult to accomplish without a large number of errors. (The same

difficulty arises in banking, where people read the amount of each check and control the printing of this number below the signature in machine-readable form). Thus, automation is still only partial.

8. Redundancy

Several times in these notes we have mentioned an important advantage of digital signals: By sending a few extra digits, we can *automatically* detect an error. With some rules for finding these extra digits, we can not only detect errors, but also automatically correct the errors.

In this section we want to look at the general subject of redundancy. How do we use extra signals to improve the accuracy of our communication?

We meet in a crowded supermarket where conversation is difficult, so I suggest that you call me when you are back home. You ask my phone number. I reply, "4236269," but the surrounding noise level is so high, neither of us is sure you have heard correctly. There are several ways we can check this:

(1) I can simply repeat the number. If you hear the same set of seven digits, the chances are you have heard correctly.

(2) You may say the number back to me. If I hear the correct number, again the odds suggest we're O.K.

(3) We may agree ahead of time to add one digit for checking purposes. For example, if we add the seven digits, we obtain 32. We can use the last digit of this sum, or 2, as a check digit. I then say to you

42362692

You know the checking rule, so you add the first seven digits and likewise obtain 2 as the check digit.

In each case, we have used *redundancy* to improve the reliability of communication. In other words, to be sure you have understood my message, we have given and received more than the basic information. In (1), we doubled the communication time by the repetition (I said the number twice); in (2), we used feedback for checking (you repeated the number so I could be sure you heard it correctly); in (3), we transmitted one additional check digit (2). If the environment is particularly noisy or if my pronunciation of the digits is not too clear, we might use even more elaborate schemes to make sure you heard the number correctly.

Redundancy in language

By nature, our language is very redundant. If we take an isolated sentence and omit 50 percent of the letters, we can usually understand the thought in the sentence. As an example, I have taken the last sentence from the following paragraph and have flipped a coin to decide which letters I will omit and which letters I will keep. If the coin came up heads, the letter was discarded; if tails, the letter retained.

National Geographic is a fine magazine -- so fine in fact, that nobody ever throws it away. Issue after issue piles up in attics and basements and barns all over America gathering dust. There is "no recycling, just the horrible and relentless accumulation of this static vehicle of our doom!" Inexorably, -h- w-i-ht -f --e --gaz-n-- --ll tr-gr-r -a-thq--k-- -- C-l--or-i, ---k c-a- min--g -owns, a-- prec-p--at- -ud s-id-s.*

The result is on the verge of readability.

From this example, we might deduce that our language is approximately 50 percent redundant: that is, information could be transmitted in half the letters. Actually, the redundancy of ordinary English is appreciably greater since a full text (rather than a single sentence) tends to repeat and re-emphasize thoughts. If we read a novel, for example, we can skip significant sections without losing the details of the plot. As you read this section, you are aware of a great deal of redundancy.

Spelled speech

When the background noise is excessive and the information must be transmitted with a high degree of accuracy, we can add enormous redundancy by switching to spelled speech: use of a word for each letter. The word "today," for example, is communicated by saying the words

Tango Oscar Delta Alpha Yankee—the first letter of each word presents the message.

The International Civil Aviation Organization has agreed on a spelled speech alphabet for use in international communication among pilots and air traffic controllers. Because the communicators are of varying nationalities and have a wide range of native languages, words must be chosen which are easily understood regardless of accent.

*The sentence reads: the weight of the magazines will trigger earthquakes in California, sink coal mining towns and precipitate mud slides. This paragraph is taken from an article in *Discover* magazine about the humor magazine for scientists, *Journal of Irreproducible Results*. *JIR* printed an article entitled, "National Geographic: the Doomsday Machine."

Alpha (Able)	Golf (George)	Mike	Tango (Tare)
Bravo (Baker)	Hotel (How)	November (Nan)	Uniform (Uncle)
Charlie	India (Item)	Oscar (Oboe)	Victor
Delta (Dog)	Juliet (Jig)	Papa (Peter)	Whiskey (William)
Echo (Easy)	Kilo (King)	Quebec (Queen)	Xray
Foxtrot (Fox)	Lima (Love)	Romeo (Roger)	Yankee (Yoke)
		Sierra (Sugar)	Zulu (Zebra)

(The words in parentheses are the code used by the U.S. military before the international agreement on the ICAO alphabet).*

Spelled speech represents an additional method of using redundancy to make sure our message is understood accurately. In contrast to the rote repetition or the feedback, spelled speech requires much more than the minimum time for communication; we effectively trade a great deal of time for vastly greater reliability.

Communicating numbers

Sending messages by the use of language can be complex because each language has many set characteristics. For example, in English:

q is always followed by u except in proper names

th starting a word is always followed by a vowel, r, or w. Likewise, if we had a phone conversation and part of one sentence was garbled, the listener could surmise what the unclear words were because of the overall content of the conversation.

To understand different redundancy schemes, we will find it simpler to talk about numerical messages — that is, a sequence of separate digits like

5 1 6 4 2 3 6 2 6 9

Here each number is an independent quantity: each digit can be anything from 0 through 9, and each is independent of the preceding or following digit.

*A fascinating (and useless unless you're a comedian) pastime is to search for a maximally incomprehensible alphabet, e.g., Knight for K, Pneumonia for P, Oedipus for O, Tsar to T, Mnemonic for M, and so on.

Let us focus primarily on two number systems: binary numbers (to the base 2) such as

0 1 1 0 1 1 0 1

where only two digits are used (0 and 1), or decimal numbers (to the base 10) such as

4 2 3 6 2 6 9

where ten digits are used (0, 1, 2, 3, 4, 5, 6, 7, 8, 9).

Number systems can, of course, have any base, although historically only a few have been common:

Base 2	Certain aborigine cultures in Africa, Australia, and South America
Base 3	One Brazilian tribe counted using the three joints of one finger
Base 4	Yuki Indians of California used the spaces between the fingers
Base 5	Most common
Base 10	Ancient China, Egypt, Greeks, Romans
Base 11	Early Maoris
Base 20	Fairly common using fingers and toes to count. Mayan system was the most advanced. It survives today in French quatre-vingts for 80. This is where the term "score," meaning 20, originated.
Base 60	Babylonians (time and angle measurement still persists)

The binary and decimal systems differ radically in the number of digits required to represent a quantity. The decimal number 115, for instance, is 1110011 in binary. In computers and communications, the increased length is offset by the advantage that only two symbols need be transmitted. If a 0 is represented by no pulse, a 1 by a pulse, the receiving equipment need only decide whether a pulse has been transmitted or not.

Transmission with error correction

We want to look for *error correction* as well as *error detection*. That is, we not only want to know when an error has been transmitted, but we also want to know what the error is. We can do this by arranging the digits in a rectangular array (with zeros added, if necessary, to complete the rectangle). For example, we want to transmit the 12-digit number

472936821743

We first arrange the digits in a 3 x 4 array or matrix

4 7 2 9

3 6 8 2

1 7 4 3

We sum each horizontal row and each vertical column and cast out 10's. (This means we would use only the last digit of the sum — i.e., a 14 would be 4.) We place these digits to the right of and below the existing numbers.

4 7 2 9	2
3 6 8 2	9
1 7 4 3	5
8 0 4 4	6

Now we add $8 + 0 + 4 + 4$, cast out tens and get 6. Then we add $2 + 9 + 5$, cast out tens and get 6. The circled 6 is the same whether calculated from the vertical column or the horizontal row.

We transmit the entire array of 20 digits in any sequence previously agreed upon with the receiver so that he can recreate the array. Any single error in the 20 digits can then be detected and automatically corrected. For example, the receiver believes the array to be

4 7 2 9	2
3 2 8 2	9 ←
1 7 4 3	5
8 0 4 4	6

↑

The receiver knows that the second row and second column do not check.

Hence, if we assume there is only one error, the entry common to this column and row must be wrong.

4 7 2 9	2
3 8 2	9
1 7 4 3	5
8 0 4 4	6

The missing entry must be a 6, determined from either the horizontal row or the vertical column.

This scheme is our first startling result of the use of redundancy: to transmit a message of 12 digits, we need send only 20 digits (12 message digits and 8 check digits) to achieve *automatic detection and correction* of any single error (or any number of errors up to three).

If we wanted to use a different method of finding these errors, we might simply send the message over and over again. We would have to transmit the 12 digits three times to get the same detection and correction as we had by using the rectangular array. This means we would have to send 36 digits (3×12) instead of 20.

While even the rectangular array is not a particularly efficient scheme (we have 8 check digits for 12 message digits), it does illustrate the capability of redundancy and gives us some appreciation of the power of the very complex redundancy in language.

Exactly the same scheme can be used with binary signals. Here if we cast out 2's, we get check digits that add up to an even number of 1's in each horizontal row and vertical column.

1 1 1 0	1
1 0 1 0	0
0 1 1 1	1
0 1 0 1	0
1 0 1 1	1
1 1 0 1	1

This form of redundancy, called parity check in computer science, is used frequently in magnetic-tape storage systems of computers.

In both the decimal and binary form, we can positively detect and correct up to three errors which occur simultaneously.

Hamming Code

The Hamming code is a more efficient use of redundancy to detect and correct any single error. The code, first presented by Richard Hamming at Bell Telephone Laboratories, is used in electronic switching systems and military computing equipment.

To explain the code, we consider a specific example in binary numbers:

1 0 1 1 0 1 0 1 1 1 0 1 0 1 0 0

This sequence of 1's and 0's reading from left to right, carries the information of the signal. Electrically, this signal can be sent as shown in Figure 12. A pulse is either transmitted or not transmitted at regular intervals of time. A pulse that is sent represents a 1, when no pulse is sent we mean a 0.

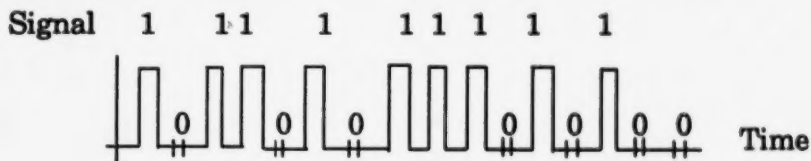


Figure 12 — Series of pulses representing the binary signal.

When noise is added during transmission, the received signal has the appearance of Figure 13. During the second pulse interval, the signal should be zero, but the noise happens to look like a pulse to the receiver. If these first six binary digits

1 0 1 1 0 1

represent the number 45 in the message, the error in the reception of the second digit means the receiver interprets the number as

1 1 1 1 0 1

or number 61 in the decimal system. (We are assuming that the above are binary numbers, converted in the usual way to their decimal equivalents.)

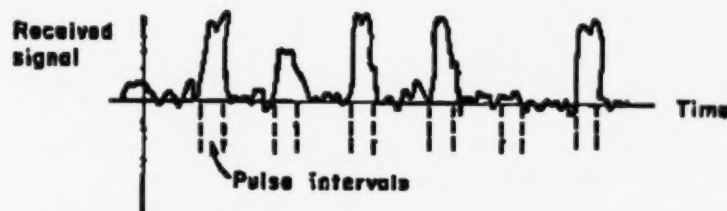


Figure 13 — Received signal when noise is added. (Only the first six pulse intervals of Figure 12 are shown.)

In order to avoid this sort of occasional error, we can introduce redundancy as follows:

1. The signal to be transmitted is first grouped in sets of four digits.

We initially consider the first four:

1 0 1 1

2. Next we draw a set of three overlapping circles called A, B, and C (Figure 14). In this figure, there are seven distinct, closed regions. Two such regions are g, common to all three circles and f, common to A and C.

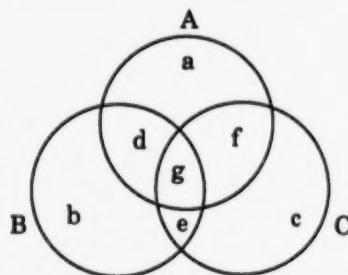


Figure 14 — Three overlapping circles with the regions a through g defined.

3. In place of d e f g we insert in order the four numbers of our message, 1 0 1 1 (Figure 15).

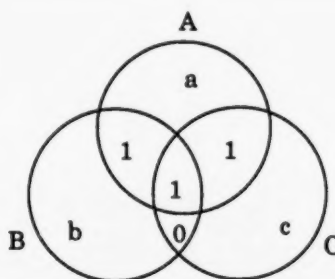


Figure 15 — The digits of the signal are inserted.

4. We now fill in the numbers for a, b, and c by choosing each as either 0 or 1 in such a way that the total number of 1's in each circle is even. In other words, in Figure 15, circle A already includes three 1's; hence, we must choose a as 1 to make the total number of 1's even (here four). Similarly, b = 0 and c = 0 (Figure 16).

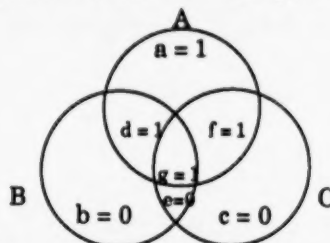


Figure 16 — Determination of transmitted signal.

5. We now transmit the seven-digit sequence,

a b c d e f g

or in our case

1 0 0	1 0 1 1
Extra digits	Original message

Each block of our digits in the original message is handled in this way. In other words, for every four message digits, we actually transmit seven digits, three more than necessary.

These three redundant or extra digits permit automatic correction at the receiver of any single error in the seven-digit block. The receiver must, of course, be familiar with the coding used at the transmitter. Then at the receiver we reconstruct the three-circle diagram of Figure 16. If any one digit is received incorrectly, one or more of the circles contains an odd number of 1's.

As a specific example, we can return to our earlier problem and assume the second digit of the message is wrong. That is, the receiver obtains

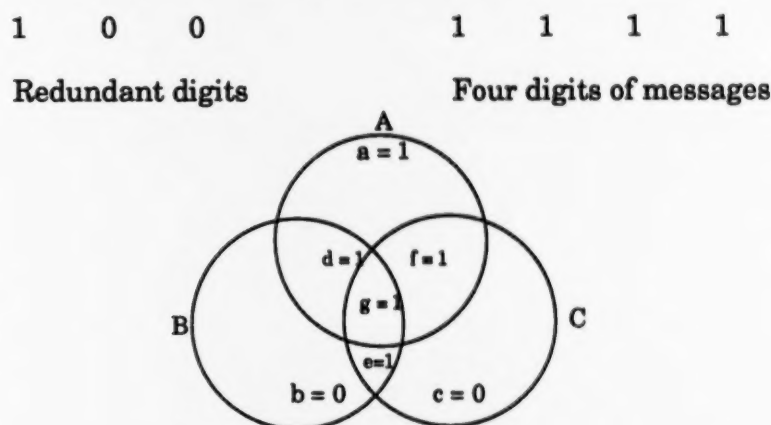


Figure 17 — Received signal including error in e.

The receiver checks the validity of this message by constructing the diagram of Figure 17.* We find that there is an odd number of 1's in circle B and also in circle C. Therefore, if there is only one error, it must be the element common to circles B

*Of course, electronic equipment does not construct a diagram. Instead, the receiver is designed to measure $a + d + f + g$, $b + d + e + g$, and $c + e + f + g$. Each of these three sums should be even. In the case of our received signal, the three sums are 4, 3, and 3: the last two are odd; the error is in the element common to B and C but not to A or element e.

and C, or element e. Hence, the received signal can be automatically corrected to

1 0 0 1 0 1 1

The communication code described above (known as the Hamming Code) uses three extra digits for each four message digits. The same procedure can be used if we draw four regions overlapping in all possible combinations (this is a complicated geometric construction.)* There are then four extra digits for each 11 message digits. Similarly, five extra digits compensate for 26 message digits. As the message block is made longer and longer, however, there is a growing possibility of two simultaneous errors: a situation which the system cannot correct.

This example is interesting because it is used in practical communication systems. It indicates the simple way in which technology uses redundancy to improve communication among men and machines. With the code described above, we can automatically correct any single error in each set of digits. With more complex codes, we can correct any one or two errors. Thus, we can build communication systems which are essentially error-free, and the task is especially simple if we use digital signals.

9. *Limits of Human Communication*

We can convert any message to digital form. In the compact disc, we store or record music as a long sequence of zeros and ones. Each second of music corresponds to about 700,000 binary digits or bits. Thus, when we are playing the disc on our home hi-fi system, the equipment is delivering about 700,000 bits/second to our ears through the sound signal.

The human being receives information through the various senses—vision, hearing, touch, taste, and smell, as well as nerves which respond to temperature, pain, muscle movement, and so forth. All these senses deliver information to the brain through the nervous system. As we sense the world around us, there is a flow of electrical pulses or signals through the nervous system to carry the information to the brain, where we make decisions on what to do and then send appropriate signals to specific muscles. In this sense, the brain is the control center accepting incoming signals and generating outgoing commands — in most cases, without conscious action on our part.

The engineer, given such a communication/control system, typically asks: How good is the system? What are the limits on what the system can do? In this example, we would ask: How much information can the human brain accept? Notice

*Because the geometry is messy, it is easier to decide on the correction digits by a systematic procedure. We visualize (but don't draw) four overlapping regions (A, B, C, and D) which overlap in all possible combinations — AB AC AD BC BD CD ABC ABD ACD BCD ABCD.

These eleven "regions" are assigned the 11 message digits. Then we determine the digits for regions — A B C D — so that each circle has an even number of ones.

that here we focus on information coming from outside the person, for example through vision, hearing, and touch, since we can then make measurements.

Vision

One way to measure information flow to a person is through the individual reading printed material. We know that each letter in our alphabet can be represented by a five-digit binary number; thus, each letter carries less than five bits of information. The average word is five letters long, so we have 5×5 or 25 bits/word in English.

As we saw in Section 8, however, the English language has an enormous amount of redundancy. At least 60 percent is redundant. So printed words carry at most 10 bits/word (40 percent of the 25 above).

A bright college student reads at a rate of about 360 words/minute or 6 words/second. Thus, in reading we accept information at a rate of about 10×6 or 60 bits/second. Advertisements for speed-reading classes always promise fantastic increases in reading rate, but 360 words/minute seems to be a good maximum figure for reading material which is reasonably unfamiliar (where we can not guess at the content of the next paragraph or page or easily skip large chunks).

There are other ways we might try to measure the rate at which human vision can receive information. For example, we can flash symbols or objects on a screen for a very short time (a fraction of a second) and test how much information the viewer has received. Such experiments confirm that the maximum information rate for vision is probably less than 100 bits/second. We obtained the number 60 bits/second from reading, but we make all sorts of assumptions — words/minute reading, bits/word in English, and so on so we really have no confidence in the figure of 60. This might well be 30 or 100. All we can say confidently is that the rate seems to be less than 100 bits/second. An average person cannot accept information at a rate faster than 100 binary digits a second. If information is sent to the person at a faster rate, the information simply does not register.

Hearing

Perhaps we do much better with hearing. One experiment involves recording speech, then speeding up (or accelerating) the speech electronically. (Cassette players with this capability are now available commercially.) Normally when we read aloud or are talking smoothly, we are speaking at a rate of about 125 words/minute. When speech is accelerated and played back to us, we can understand it only up to a rate of about 300 words/minute; at higher rates, we are lost.

According to our calculations under vision, 360 words/minute corresponds to about 60 bits/second. Consequently, the maximum rate for listening to speech (300 words/minute) means the listener is accepting information at about 50 bits/second

— certainly under our estimate of 100 bits/second as the maximum human rate.

Touch

Tactile communication seems to give a similar result. Here we can touch a person briefly in one area for a one, in another area for a zero, as we discussed in Section 1. We might also try a different scheme using six different areas and the Braille code or we might go to even more locations.

As we increase the number of areas, learning becomes more difficult for the person receiving the information. But experiments have indicated that no matter what scheme we use, we cannot achieve information rates higher than about 50 bits/second.

Combining senses

The military has been interested in maximizing human performance and has actually conducted experiments using both vision and hearing senses simultaneously. There seems to be no evidence that the maximum information rate can be appreciably raised by such a combination.

Thus, experimental evidence acquired so far seems to indicate that the human being is able to accept information only at a rate of less than 100 bits/second. The experiments mentioned above are, of course, highly restricted in the types of signals given to vision or other senses. Perhaps when we look at a new scene (e.g., on a slide) for a second, we do absorb more than 100 bits of information. Even the very exploratory research on vision, however, does seem to confirm that there is a rather severe limit on human capabilities, perhaps a limit near 100 bits/second.

If we accept such a limit, we then might ask: Why are we recording over 700,000 bits/second on a compact disc—7,000 times as much as any listener can absorb? The situation is even more extreme in television. While we have not discussed TV in this module, it can be shown that a TV video signal is capable of carrying 30,000,000 bits/second of information.

Clearly, when we are communicating with people electronically, our systems are using "technological overkill." The technology is enormously more elaborate than necessary. The reason is simple: The engineer or designer does not know what information the human vision or hearing needs to be given for complete understanding of the message. What do we actually hear as we listen to recorded music, see as we watch TV? Unable to answer these questions, the engineer simply builds the audio or video technology to transmit everything. Human physiology then can select the parts of the sound or picture it can use.

Perhaps in the 21st century we will understand human senses and brain operation and then be able to design systems that match our capabilities.

APPENDIX A

NYQUIST SAMPLING THEOREM

Hi-fi enthusiasts know that the signal for quality music contains a wide range of frequencies (signals of different pitch) — from 30 Hertz or cycles/second to 18,000 Hertz (from three octaves below middle C to six octaves above middle C). This is the range over which people can hear. Indeed, a sound system which is “high-fidelity” must faithfully reproduce sounds over this entire range of frequencies.

When we study signals of all sorts, we find that any signal can be thought of as the sum of pure notes at different frequencies.* A long-distance phone conversation has all frequencies up to 4,000 Hertz (phone conversations are not high-fidelity). The motion of a boat buffeted by waves may have frequencies up to 2 Hertz. Every signal has a maximum frequency — the more rapidly the signal changes, the higher the maximum frequency.

The Nyquist sampling theorem states that no information is lost in the sampling process if samples are taken at a rate higher than twice the maximum frequency within the signal.

Thus, the speech signal which the telephone company sends in a long-distance call by satellite or cable has no frequency components above 4,000 Hz. Before my voice signal is transmitted, the phone company throws away all the components of high pitch above 4,000 Hz, which is four octaves above middle C or about the highest note on the piano. This is one reason your voice on the phone sounds slightly different from your voice in person; you are perfectly understandable over the phone, but your voice signal is missing the musical, high-frequency parts.

Since the telephone company transmits only voice components below 4,000 Hz, the signal can be exactly represented by samples taken 8,000 times a second. To send your speech signal from Los Angeles to Boston, the phone company need only send 8,000 numbers each second. This may not seem like a major advantage, but typically each number can be sent in much less than 1/8000 second — consequently, they send one number, then there is a time interval when the equipment is not used while they wait for the next sample 1/8000 second later. During this unused inter-

*In mathematics, this statement is known as the Fourier theorem. Jean Baptiste Joseph Fourier (1768-1830), orphaned at eight, was a child genius. By the age of 14 he was ghost-writing sermons for one of France's leading clergymen. At 16 he was a math teacher. In 1798 Napoleon took him on his Egyptian campaign as the first science advisor to a ruler. The role of science advisor was established at the outset: when Napoleon rushed home to reorganize his army, he forgot Fourier, who remained in Egypt for several years. His experience there convinced him that desert heat was healthy; in later years, he always dressed in heavy clothes and kept his Paris apartment at 90°F. Perhaps the fact that visitors seldom stayed very long accounts for his great scientific productivity, primarily in the topic of heat flow.

val, they can send samples from other phone conversations. Thus, the single set of transmission equipment can transmit many phone conversations concurrently.

This technique for utilizing long-distance equipment is called *time multiplexing*. Multiplexing is sending several messages at once; *time* multiplexing means we share the available time (Figure A1). This sharing of expensive equipment means that the cost of long-distance calls is a small fraction of what it would be if the complete system could handle only a single call.

This same time sharing has now become standard practice in the space program (where the Houston control center wants to monitor many signals describing the conditions of the astronauts and vehicle) and in an automotive assembly plant (where the computer scheduling work or parts delivery must know what is going on at each work station along the assembly line). When the signals appear as samples, time sharing is possible.

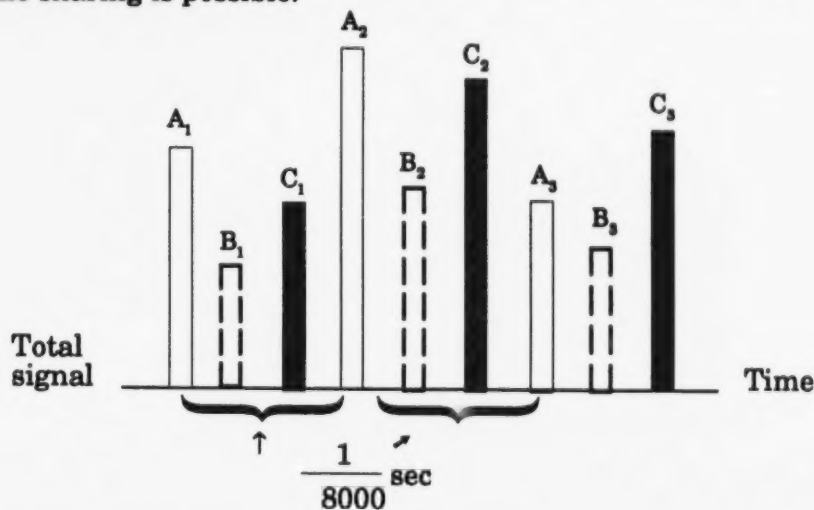


Figure A1—Time multiplexing of three phone signals: A, B, and C. Each signal is represented by samples appearing 8000 times/second. Sample A_1 , A_2 and A_3 are for conversation A. In an actual phone system, hundreds of conversations travel concurrently through the same equipment, but here we show only three for simplicity.

APPENDIX B

TELEPHONE—AN ANALOG SIGNAL

In 1876, telegrams were a limited and impersonal way to communicate. Then Alexander Graham Bell patented the telephone, which permitted direct electronic communication, person-to-person. Once again the new technology became commonplace with astonishing rapidity.

Initially, appropriate uses for the telephone were not clear. Businesses in New York City communicated quite adequately through messenger boys on bicycle or foot. I have always thought the greatest salesmen were those who sold the first phone in a town to convince someone to buy a phone when there was no one else to call! In London and Paris, the phone company placed transmitters in theaters; subscribers could listen to plays from their home. Thomas Edison invented the phonograph in the search for a device to record telephone messages (no one visualized both parties available at the same time). The telegraph industry tried unsuccessfully to require that phone subscribers pay for each word spoken.

Then doctors and lawyers became subscribers, and the technological revolution was underway. By 1900, only 24 years after the invention, the telephone was commonplace in the major cities, and 50 years after the invention more than half of American homes had the technology. The impact in New York City, for example, was enormous. The messenger boys disappeared. Skyscrapers were now possible (without the telephone, there would have been no chance of enough elevators to handle the messenger boys essential for the conduct of business).*

The telephone moved us technologically away from digital signals. A telephone message was an electrical reproduction of the sound signal coming from the speaker's mouth. In the wires connecting speaker to listener, the current varied in the same way as the air pressure at the speaker's mouth. We lost all the advantages of the digital signal.

Today, however, even speech or music signals are communicated or recorded in digital form.

*Engineers like to argue about which technology was most important in allowing skyscrapers and the growth of the central business districts in cities: the elevator or the telephone. The elevator is an astonishing technology, with a safety record for vertical travel far surpassing any technology of horizontal travel.

APPENDIX C

DIGITAL AUDIO OR COMPACT DISCS

In the last three years, audio hi-fi equipment has gone digital. The compact disc (CD) or digital audio provides unusually pleasing music, according to many hi-fi enthusiasts, and the CD threatens to replace the long-playing, 33 1/3-rpm record.

The compact disc works as follows: When the disc is made, the manufacturer first converts the audio or music signal to digital form — a long sequence of binary digits. These zeros and ones are then recorded on circular paths around the disc. A path consists of occasional raised portions (Figure C1). In each time interval, a *change* in height means a one is recorded, *no change* represents a zero.

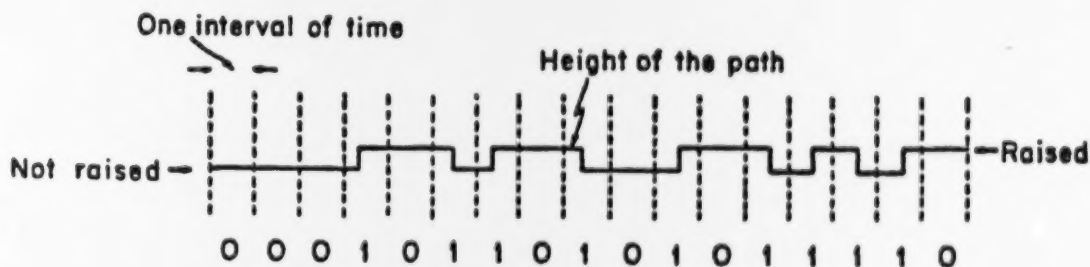


Figure C1 — Side view of one part of a circular path around the disc. Part of the path is raised, part is not. In each interval, no change represents a zero, either change (up or down) represents a one.

The technology of making records is awesome. Each change in height is only one sixth of a millionth of a meter.

Once the music is recorded, how do we recover it? As the disc spins, a laser beam (a sharply focused light) shines on the path moving under the beam. If there is no change in the height during an interval, the laser beam reflects off the disc to a light sensor and the equipment reads a zero. If there is a height change, the light is diffused and reflects in different directions, the sensor receives relatively little light, and the equipment reads a one for that interval.

Thus, there is no mechanical contact, no needle riding on the disc as in the older hi-fi systems. The only contact with the disc is by a light beam which exerts essentially no pressure to cause wear or deterioration of the disc with repeated playing. The compact disc has other remarkable advantages over the older 33 1/3-rpm disc, as we see in the following paragraphs.

Complexity of the digital signal

The best hi-fi systems reproduce all frequencies from about 20 Hertz (cycles/second) out to 18,000 Hertz — from 3 1/2 octaves below middle C to 6 octaves above middle C. This is roughly the range of human hearing (Figure C2); hi-fi or high fidelity means we must reproduce faithfully all the sounds people can hear. Thus, there is really no reason for an audio system to record orchestra sounds at frequencies higher than about 20,000 Hertz, even though a violin (for example) does give sounds up to 60,000 Hertz.

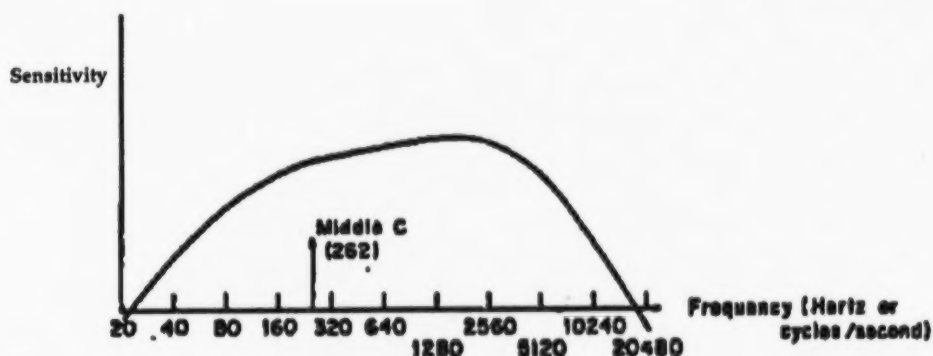


Figure C2 — Sensitivity of human hearing for a healthy young person. Hearing is most sensitive (that is, can detect the weakest signal) when the frequency is about 2,000 Hertz. 18,000 is about the maximum frequency, although infants and a few adults go to higher frequencies.

In recording music on a compact disc, the recording company must sample the sound signal on the way to generating the digital signal. This first step in going to a digital signal is done with 44,000 samples/second. By the Nyquist sampling theorem (Appendix A), the digital signal can therefore preserve all sound frequencies up to 22,000 Hertz (one half the sampling rate), or well beyond the limit of human hearing.

The size of each of these samples then must be described by a binary number (the quantizing and coding steps we discussed previously). How many binary digits should be used? In Section 5, we used five bits (binary digits) to describe body temperature. This is nowhere near enough for high-quality music, however, and the compact disc uses 16 binary digits for each sample.

Now we begin to have a picture of the digital signal. We have 44,000 samples every second; each sample is described by 16 binary digits. Thus, we have 44,000 x 16 or 700,000 bits/second in the digital signal which represents the music. Each second of sound has to be represented by more than 700,000 zeros or ones.

AUDIO RECORDING

- 1877 Edison invented a sound-reproducing system: to record, a needle dented tin foil. Edison envisioned the principal use would be recording incoming phone calls (the answering machine).
- 1855 Chichester Bell and Tainter used a cardboard cylinder coated with wax.
- 1887 Emile Berliner invented the gramophone, the first flat disc.
- 1890 Poulsen, Denmark, demonstrated the first magnetic recording of sound along a steel wire. The system was never used commercially.
- 1925 Electric-motor drive replaced the manual wind-up.
- 1939 German engineers used plastic tape coated with magnetic material, for the first useful magnetic recording.
- 1948 Peter Goldmark and CBS Laboratories introduced the long-playing high-fidelity, 33 1/3 rpm disc. A 12" disc held one half hour of music.
- 1950 Tape replaced records in radio stations.
- 1955 Stereo tape recorders appeared, along with videotape.
- 1958 Stereo records appeared, with the monaural signal, the radial motion of the pick-up, and the stereo information simultaneous up-and-down motion.
- 1961 Audio cassettes were sold in the Netherlands, then came to the U.S. in 1964.
- 1965 Cartridge recorders were manufactured for cars.
- 1968 Dolby system was introduced to reduce the hiss made by the tape at all times.
- 1983 Compact discs appeared.

The usual disc holds 74 minutes of music. If we perform the simple multiplications, we find that each disc holds more than 3,100,000,000 bits.

This number is so large that it really has very little meaning for most of us. Figure C3 illustrates how densely the information is packed. The large circle represents a typical disc. Within the small circle, there are 900,000 binary digits stored. In this region, the system must decide 900,000 times whether there has been a change in level or not. Obviously, the laser beam must be aimed at precisely the correct spot on the disc. This technological accomplishment is even more astonishing when we realize that the system works in an automobile moving over bumpy streets.

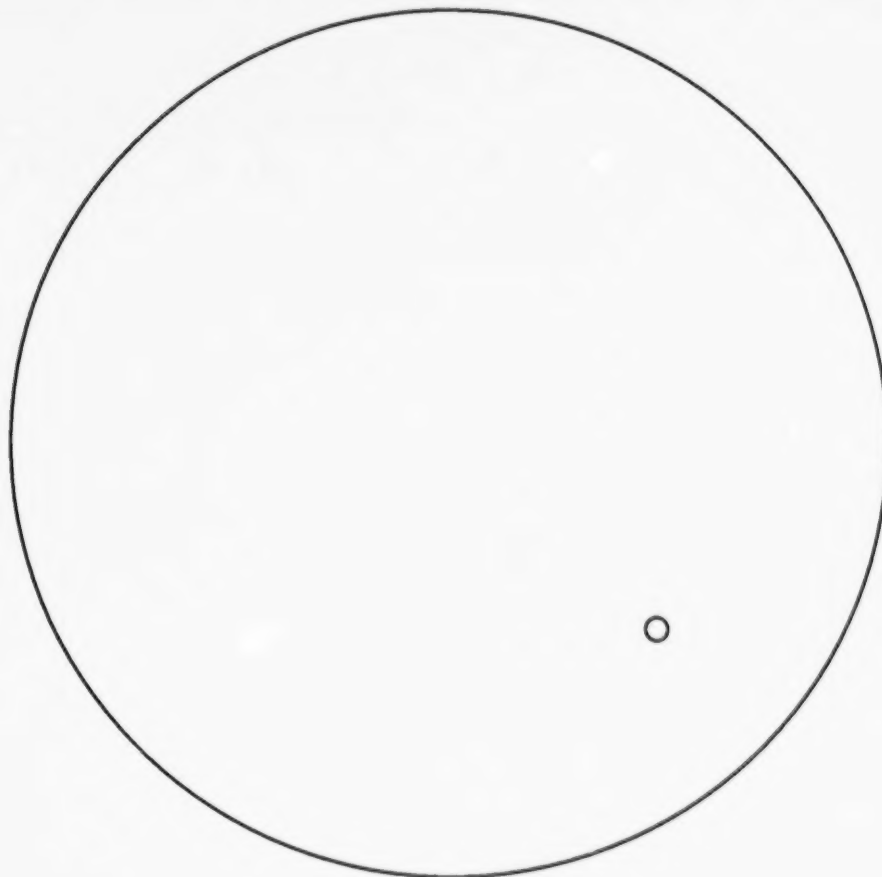


Figure C3 — The large circle is the size of a compact disc. To illustrate how information is packed densely, the small circle holds about 900,000 bits of information.

Significance of this complexity

The choices of 44,000 samples/second and 16 bits/sample mean that the CD gives major improvements in the performance of hi-fi-systems. Indeed, these numbers are so large, we might ask why were they chosen? Both numbers exceed the capability of human hearing. Since hardly anyone can hear frequencies above 16,000 Hertz, why not sample at 32,000 per second rather than 44,000? Similarly, 12 bits/sample are about as much detail as human hearing can distinguish, so why go to 16? If the manufacturers had used 32,000 and 12, the recording and playback equipment would have been much simpler.

There are arguments for the higher performance of the CD. These digital re-

cordings will last forever and hence represent permanent records of today's music. We should anticipate future demands and then design technology, when possible, to meet those needs. For the first time, we have the technology to do a superb job of recording sound; if we do not make the best equipment possible now, will it be obsolete in 60 years?

There is also a strictly commercial argument if one assumes industry is motivated only by the "bottom line." To convince customers to discard their conventional turntables and 33 1/3-rpm discs in favor of digital audio, the manufacturers must offer a product unquestionably better. Around 1950, when CBS introduced the 33 1/3-rpm record, audio enthusiasts did discard the obviously inferior 78-rpm equipment and make the major investment required for the change-over. Quadraphonic sound (three-dimensional instead of the two-dimensional stereo) was, in contrast, a marketplace failure, as the public never considered the music quality significantly better.

Advantages of the CD system

The question arises: is the CD (compact disc) system really better than the older, 33 1/3-rpm system? Two years ago, when the digital audio equipment first appeared, engineers at California Institute of Technology carried out an experiment to answer this question.

They invited manufacturers of both analog (33 1/3-rpm) and digital (CD) equipment to set up systems, then asked trained musicians to compare the performance without being aware of which system they were listening to at any time. The results were inconclusive: neither system was significantly better in the opinion of these musicians.

Such an experiment is interesting, but not really too meaningful in anticipating customer acceptance. Most hi-fi enthusiasts do not have unlimited money to spend on equipment, acoustically ideal rooms in which to locate speakers, or the resources to use only new discs. Consumer acceptance of digital equipment over the past three years indicates that many buyers *do* find the new equipment significantly more pleasurable.

What are the principal advantages of digital audio? We can group these in two categories. First, we have the potential capabilities for accurate reproduction of sound:

- (1) As noted above, the digital system can handle audio frequencies out to

*Actually, some manufacturers are considering audio discs which record all frequencies out to 60,000 Hertz or more. Human hearing goes only to 18,000 at best. However, when there are sounds within the music at much higher frequencies, there can be beat notes within the audible range. For example, a note at 36,000 will beat with one at 40,000 to give a 4,000 Hertz signal which we can hear. When we listen

22,000 Hertz. This is a slight improvement in fidelity when compared to the best analog system.*

Furthermore, the high-frequency performance of the 33 1/3-rpm disc deteriorates rapidly as the pick-up moves in toward the center. The first song recorded may carry frequencies out to 18,000 Hz, but the last song has less than half this frequency range. The problem is that the disc turns at constant rotational speed (33 1/3-revolutions per minute); near the end of the recording, the distance the needle travels in one revolution is small. Since the signal is the speed of the stylus, the grooves must have wider excursions to offset their lower speed in order to maintain signal fidelity.

The digital system avoids this problem by increasing the rotational speed of the turntable as the laser pick-up moves inward, so that the number of bits per inch is about the same throughout the recording. Every song appears with the same fidelity.

(2) The digital system uses 16 bits to describe each sample. Sixteen doublings of size are possible as we go from the weakest to the strongest signal. In terms used by audio engineers, each doubling corresponds to an increase in sound amplitude of six deciBels (6 dB).** Thus, using 16 bits means that the strongest signal can be 16×6 or 96 dB louder than the weakest signal. In the terminology of the hi-fi enthusiast, the "dynamic range" is 96 dB for digital audio.

This capability contrasts markedly with performance of analog systems. The 33 1/3-rpm record gives a dynamic range of about 66 dB, limited by the maximum allowable radial motion of the pickup because of its mechanical nature (in particular, its inertia). Audio tape has a slightly larger dynamic range of about 76 dB, still much less than the CD's 96 dB.

Furthermore, in both the mechanical and magnetic analog systems, there is a certain amount of noise present even with no signal. In the mechanical system, the needle moves slightly radially because of its natural oscillation and imperfections in the circular groove. In the magnetic (cassette) system, there is always some residual magnetism.

As a consequence of these features, the digital system is remarkably quiet during pauses in the music, then can move from very soft to unusually loud sound.

The second group of advantages of the digital system arise from the flexibility of digital signals and the nature of the play-back mechanism:

to an orchestra directly without recording of the music we do hear such beat notes, although they are very small and usually not noticed. A symphony orchestra in a music hall does sound different from even a CD recording to individuals who have exceptional hearing capability.

**This is really the definition of dB or deciBels.

(1) There is essentially no contact with the disc, so no wear. Hi-fi enthusiasts often find that quality deteriorates seriously after a few playings in a 33 -1/3-rpm system.

(2) Having the signals in digital form (a sequence of zeros and ones) allows minor imperfections of the disc to have no effect on the sound: the pick-up only needs to be able to determine in any interval whether there is a change in height or not.

(3) The disc will play properly even if there is a small hole at one point (even a hole the size of the small circle in Figure C3). Automatic correction for these errors comes from two techniques:

(a) Successive binary digits are recorded on different parts of a rotation, rather than together. Consequently, a hole or defect in the disc affects a set of numbers drawn from many *different* samplings of the original sound signal.

(b) After the special-purpose computer in the playback system restores the binary digits to their proper sequence, redundancy is used for automatic error checking and correcting. (We saw one example of error detection in Section 1.)

Thus, the playback computer automatically corrects any errors introduced by the small disc defect.

To conclude this section on digital audio, we should emphasize that the CD system possesses its various advantages because it uses digital signals. Once we convert the sound signal to a sequence of zeros and ones, we can use all the power of electronic digital techniques (including special-purpose computers) to process the signal to record, pick-up, and assure fidelity. Even though the number of bits required each second seems overwhelming, we gain all the important advantages inherent in digital signals.

PROBLEM 1

EXPONENTIAL GROWTH

The introduction points out that Figure 1 shows exponential growth. By this term, we mean that, for each unit of time, the quantity grows by the same multiplying factor. In our case, each 34 years the communication capacity multiplies by 100.

In this specific figure, we see that

Year	Capacity
1864	1/100
1896	1
1930	100
1964	10,000

If growth had continued at the same exponential rate, in what year would the capacity have reached 10 billion phone conversations on one system — in other words, enough for everyone in the world talking simultaneously on the one system, with some capacity left over for TV signals?

This growth is called “exponential” because the mathematical equation involves the exponential we met in algebra. The equation for the straight line portion of Figure 1 is

$$\text{Capacity} = \frac{\text{Year} - 1896}{(100)}$$

PROBLEM 2**SETI**

Below we describe the digital radio message sent into space in 1974. From this explanatory material, answer the following questions.

- (a) Figure P2-1 shows the sequence of zeros and ones broadcast by radio. Explain exactly how this sequence corresponds to the left half of Figure P2-2.
- (b) The message is in digital form. What arguments can be given for using digital rather than decimal form?
- (c) Some distinguished scientists and engineers have argued that we should not be sending out messages which reveal our location and level of intelligence. Alien civilizations are apt to be much more advanced and may come to attack us. Other scientists have argued against this concern by pointing out that we have already sent out signals now up to 60 light years away, and we certainly can not hope to keep our existence a secret. Write a paragraph for whichever side you take in this debate.

THE ARECIBO MESSAGE

In 1974 work was completed on both a reflector and transmitter for the giant 1000-foot radio telescope at the Arecibo Observatory in Puerto Rico. When the half-million watt power from the transmitter was focused by the telescope's reflector, it could send forth the strongest signal ever to leave Earth up to that time. Frank Drake writes that "the signal was perhaps a million times brighter than the Sun at a comparable wavelength, a signal (that would be) easily detectable by radio telescopes no more sensitive than ours, across interstellar distances of many thousands of light-years." It was decided that a message to the stars would be composed and transmitted on the occasion of the dedication of the Arecibo radio telescope.

The following description of the message was written by Frank Drake.

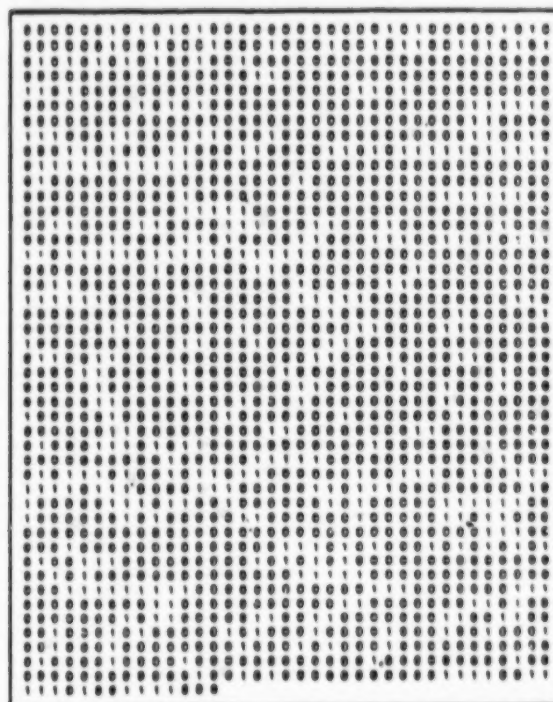
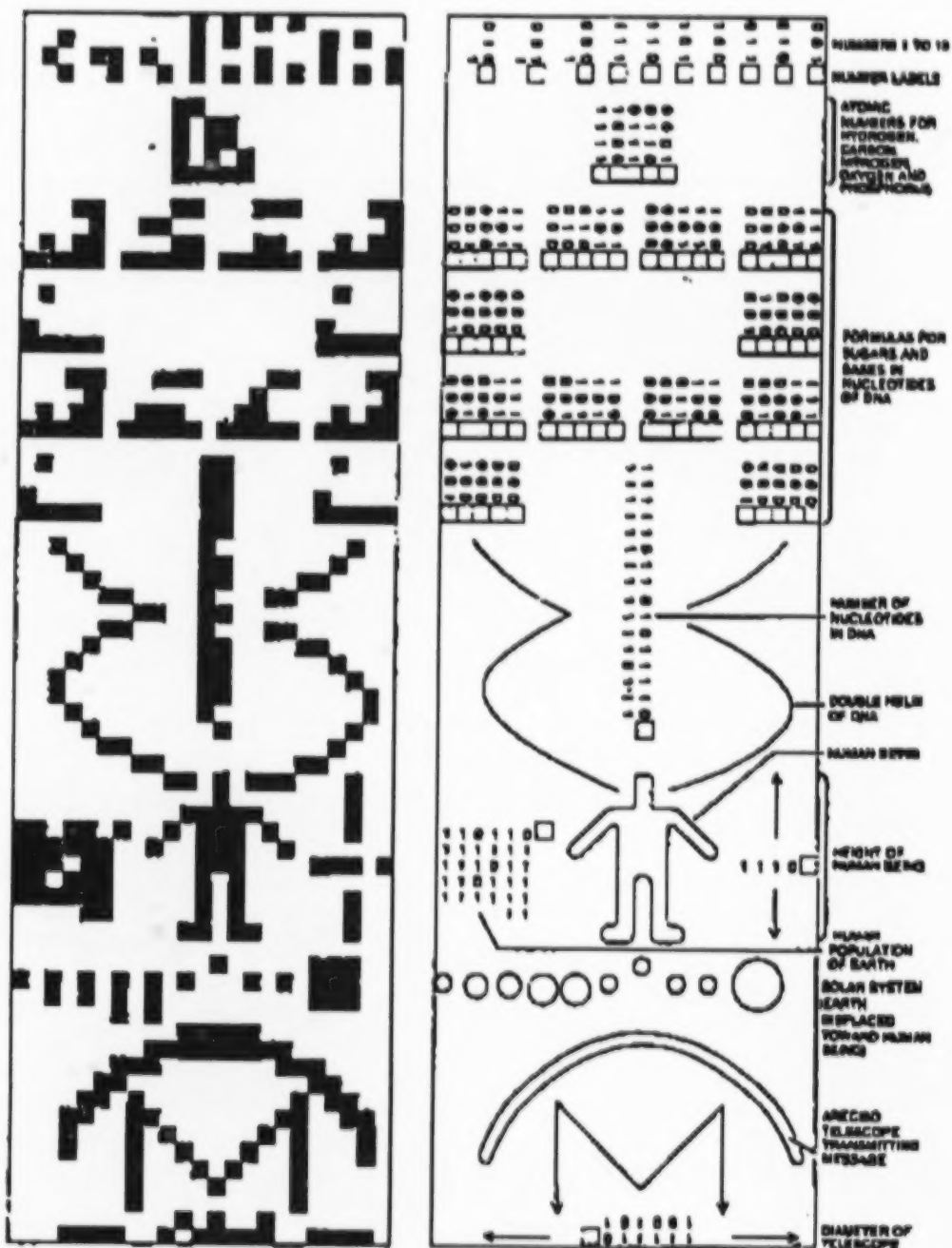


Figure P2-1
Digital message
transmitted.

ARECIBO MESSAGE IN BINARY CODE was transmitted in 1974 toward the Great Cluster in Hercules from the 1,000-foot antenna at Arecibo. The message is decoded by breaking up the characters into 73 consecutive groups of 23 characters each and arranging the groups in sequence one under the other, reading right from left and then top to bottom. The result is a visual message that can be more easily interpreted by making each 0 of binary code represent a white square and each 1 a black square.

The message starts with the numbers 1 through 10, again written in binary code, but intentionally squeezed against the top margin so that we were forced to adopt some convention to indicate how large numbers are to be written when there is not enough space to do. In this way, the convention would be clearly established. This has to be done carefully and clearly because numbers appear in the message, written not only on several lines but in three orientations. Then follow the numbers 1, 6, 7, 8, and 15, placed in prominent isolation. This is probably the most enigmatic part of the message. The only self-consistent interpretation of these numbers is that they are the atomic numbers of the elements hydrogen, carbon, nitrogen, oxygen, and phosphorus. This makes sense particularly when one explores the next characters in the message, which are sets of numbers in the same locations as the atomic numbers of the various elements. These make up chemical formulas—in fact the formulas of the components of the molecule that determines the nature of terrestrial life, the DNA molecule. Its composition is shown here and even its form, that of a double helix. When one looks at the usual sketches of the complicated DNA molecule, it may not seem possible to depict it clearly and succinctly with the



From "The Search for Extraterrestrial Intelligence" by Carl Sagan and Frank Drake. Copyright © 1978 by Scientific American, Inc. All rights reserved.

limited imaging capabilities of this type of message, yet when we set out to do it, it turned out to be rather simple after all.

The double helix winds into the head of a human being, indicating a relationship with the intelligent creature. Various attempts were made to make a better unisex version of the human, but the next-best version looked as much like a gorilla as a human, and so I had to settle for this rather masculine-appearing person. That the DNA is important to us is clear. Of equal importance, a large number placed within the DNA molecule tells of the number of nucleotide pairs, or code bits, in the typical human DNA molecule. This expressed something that couldn't be described—otherwise in such simple artwork—the level of our evolution, and some measure of the level of our intelligence.

To the creature's right is a measure of its size, given in terms of the wavelength on which the message was transmitted, 12.6 centimeters. To the creature's left is a large number giving the population of humans on earth. Below the human is a sketch of the solar system, including the sun and all nine planets, and showing at least the relative sizes of the planets. Planet 3 is offset from the rest toward the human, showing that there is something special about it, that it is, in fact, the home of the human race. Finally, below the solar system and "up" with respect to the Earth is a sketch of a telescope, a reflector focusing rays to a point. Just below this is a number giving the size of the telescope, about 1,000 feet, again in wavelengths. This is both the size of the telescope that sent the message, the Arecibo radio telescope, and the size of the largest radio telescope on Earth. Thus, we described the state of advancement of our technology. Note that if the message is sent over and over, it will seem to flow out of the telescope, making the point that this is the one that is sending the message.

The message was successfully sent on November 16, 1974, at 1:00 AST (Atlantic standard time) as approximately 200 people watched. It was directed toward the great globular cluster of stars in the constellation Hercules known as Messier 13, some 25,000 light-years away. It will take the message 25,000 years to reach the 300,000 stars of Messier 13, passing perhaps 30 other stars along the way. As the audience looked at the huge transmitting structure 500 feet above them, shimmering in the tropical sun, many sensed that there was something very special about the eerie whine, the sound of a message being transmitted to the stars for the first time. It took 169 seconds to send, and as the warbling of the message changed to the steady tone that marked the end of the message, the emotional impact on many of the audience was evident—there were tears in many eyes and sighs to be heard. Brighter than the fires of our own sun, the message was on its way. The first bits of the message were already passing the orbit of Mars. Only seven hours later the message had passed the orbit of Pluto and was plunging at the speed of light into the darkness of the interstellar space. It is now farther from the Earth than the distance to the nearest star.

PROBLEM 3

The BART mass transportation system in the San Francisco area has been beclouded by arguments and public criticism because of large cost overruns, unexpected delays, early safety problems and difficulties with the automated train control system, low ridership, and failure to alleviate traffic problems. Now that several years of experience are available, there is still a strong feeling that the project was foolish.

Much of the technology is interesting. As one very small example, the central computer sends out electrical signals to each train to indicate the desired speed. Eight different speeds can be ordered, each represented by a six-digit binary signal:

130 Km/h	101111
113	100111
80	101011
58	100011
44	100101
29	101001
10	100001
0	100000

Each order can be sent in $1/3$ second, each bit of information (a zero or a one) in $1/18$ second, and the order is repeated continually.

(a) To send eight different speed commands, we only require three bits or binary digits. When we chose a signal of six binary digits, we obviously have available an enormous amount of redundancy. Recognizing that the receiver of this signal may not know *when a particular signal begins and ends*, describe briefly how this redundancy is used.

(b) Can the same operating characteristics (i.e., successful communication without a synchronizing signal) we obtained with five-digit signals — in other words, are six digits needed?

Hint: While we might look for a general formula for the number of digits needed, it is easier to look at each case. For example, if we use a signal of four binary digits, the possibilities are:

A	0000	I	1000
B	0001	J	1001
C	0010	K	1010
D	0011	L	1011
E	0100	M	1100
F	0101	N	1101
G	0110	O	1110
H	0111	P	1111

But signals with only one 1 are indistinguishable when a repeated signal is sent. For example,

BBBB	0001000100010001
CCCC	0010001000100010

If we started observing at an unknown time within this sequence, we could not tell which signal we are picking up. Hence, if we use B, we can eliminate C, E, and I. We are left with

A	0000	J	1001
B	0001	K	1010
D	0011	L	1011
F	0101	M	1100
G	0110	N	1101
H	0111	O	1110
		P	1111

Similarly, only one of the signals with two 1's in a row and two 0's can be used. If we use D, we eliminate G, J, and M. Three 1's in a row uses H, but eliminates L, N, and O. We now have:

A	0000	H	0111
B	0001	K	1010
D	0011	P	1111
		F	0101

Finally, if F is used, K can not be, and we have only six possibilities:

A	0000	F	0101
B	0001	H	0111
D	0011	P	1111

Four digits are clearly not enough to send eight speeds if we insist that each speed is sent several times and reception can begin at any time in the sequence.

PROBLEM 4

GRAY CODE

The Gray code is shown in Section 6 (p. 19). This code has the interesting property that only one digit changes as we go from one number to the next larger (or smaller).

By inspection of the example given in Section 6, determine the rule or algorithm for generating the sequence of binary numbers. Apply this rule to list in sequence the five-digit binary numbers representing the decimal numbers 0 through 31.

PROBLEM 5

TACTILE COMMUNICATION

In comparison with the eye, the ear is exceedingly poor in detecting the direction from which a signal comes, but much better in noticing every short change in sound. For example, the ear can detect a sound break only $1/500$ of a second long. In this sense of determining brief signal changes, the sense of touch performs between hearing and vision. The human being is able to detect an interruption of only $1/100$ second in a steady pressure applied to the skin.

Thus, we might communicate with a human being by attaching to his skin buzzers or pressure devices at perhaps a dozen points around the body (points well separated and where the skin is sensitive to pressure).

By applying different signals at the various locations, we can develop a complex code for sending information to the human being.

(a) We desire to calculate the rate at which information can be communicated to a person through a system of vibrators attached to the body at six different locations. Each signal is a pulse lasting 0.1 second. How many different symbols do we have if we can vibrate any one point, any pair of points, any triplet, and so forth, up to and including all six at once? (Note the similarity between this system and Braille.) Actually, we have to choose the sites with some care. They must be far apart, not connected to the ear (otherwise you "hear" the vibration) and not on corresponding points on the two halves of the body (why?).

(b) We might increase the number of different signals by making each signal one of two possible intensities (e.g., soft or strong). Actually, one can detect more than a dozen different intensities of pressure or vibration at a single point, but in rapid communication we want to use only signals which can be easy and reliably detected.

With this greater flexibility, how many different signals are now available? In the next paragraph, we call this number N .

(c) There are now N different signals which we can deliver to the man in any $1/10$ th of a second. The maximum amount of information which can be conveyed in $1/10$ second is then given by the equation

$$2^x = N$$

where x is the amount of information in bits. The equation says that the number 2 raised to the x power is equal to N . We can find this amount of information without using the equation at all if we are satisfied with an approximate answer. We simply look at the table shown below.

x bits	2^x	x bits	2^x
1	2	7	128
2	4	8	256
3	8	9	512
4	16	10	1024
5	32	11	2048
6	64	12	4096

If N is 400, we see the value of x lies somewhere between 8 and 9 (since 8 corresponds to $N = 256$ and 9 to $N = 512$). Hence, if there are 400 different signals, we can send a little more than 8 bits of information. How many bits are represented by each of our N signals?

(d) For our value of N , how much information can be sent every 0.1 second? How much information per second? How does this compare with information sent through the eye in reading? To build a system such as this, we would have to train a person to recognize or understand the incoming signals. Does such a scheme seem feasible?

Actually, this idea of communicating through the sense of touch is hardly new since it was suggested by Jean-Jacques Rousseau in 1762 in his treatise "Emile." It is only in the last few years, however, that we have begun to understand scientifically the types of touch signals which can be understood by the human, the structure of communication signals, and the ways in which signals might be coded and presented to a person for easy and reliable understanding. Experimental systems

have been built which give rates of communication appreciably above those of telegraph transmission, but we still are a distance from matching the rate of communication which can be achieved by speech. Interest in this research comes from the military, anxious to communicate in dark and noisy environments, and from the problems of communicating with blind or with blind and deaf people.

OPTIONAL SUPPLEMENT

We might also try to obtain some measure of one's information capabilities through the nerves leading from the hand to the brain by considering the activities of a good pianist. If each of the 10 fingers can hit any one of three notes, how many different combinations are there (in other words, how many different signals)? Actually, nowhere near all these combinations are equally likely, so we might estimate a much smaller number as the number of different signals. If the pianist plays at a rate of 300 notes/minute, what is the corresponding ability of the nerve system to carry information, measured in bits/second?

PROBLEM 6

UPC

The term "supermarket automation" refers to the use of the Universal Product Code (UPC) at the check-out counter. Each item carries a bar code (Figure P6-1) which can be read by a laser beam (which reflects off the white area, is absorbed by the black). In a supermarket each grocery item has a 12-digit label such as

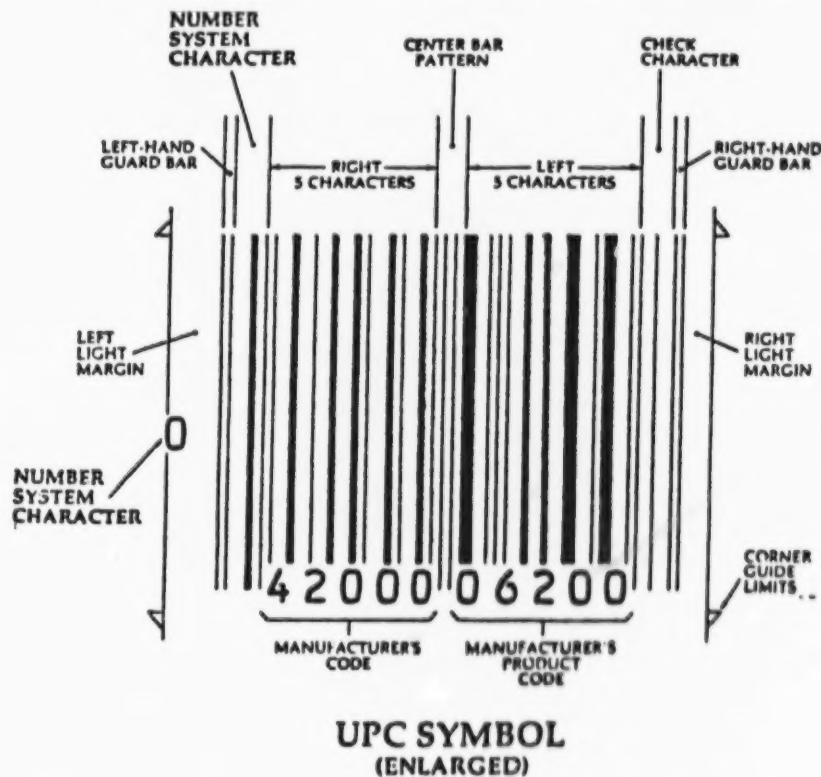


Figure P6-1— Example of Universal Product Code

- Here
- (1) The initial zero means this is a grocery product
 - (2) The next five numbers identify the manufacturer
 - (3) The next five numbers identify the particular item
 - (4) The last number is a check digit to verify that the device has read the code correctly.

In the bar code, each decimal digit (of the 12 needed) is represented by seven

binary digits. In the left half the code is:

0 - 0001101	5 - 0110001
1 - 0011001	6 - 0101111
2 - 0010011	7 - 0111011
3 - 0111101	8 - 0110111
4 - 0100011	9 - 0001011

In the right half, the complementary code is used (each 0 above replaced by a 1, each 1 by a 0). Thus, in Figure P6-2 we can identify each seven binary digit with the corresponding decimal identification number.

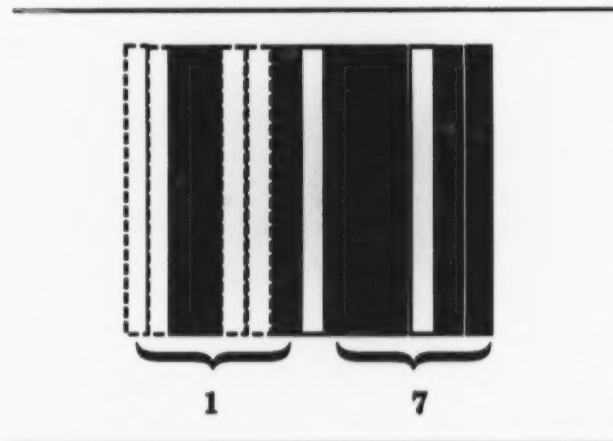


Figure P6-2 — Bar representation of 17 on the left side.

This code uses a large amount of redundancy (seven binary digits means 128 possibilities and we only need 10). Each kind of redundancy is inserted to simplify the automatic reading equipment.

(a) Each seven-digit code on the left has an odd number of one's, on the right an even number. What flexibility do we gain by this?

(b) Each seven-digit package starts with a zero and ends with a one if we read left to right. How does this feature simplify reading?

Each seven-digit packet has exactly three changes from zero to one. Thus, 0000111 is not used since there is only one change, 0101011 is not used because there are five changes. Why does this feature improve reliability of reading?

(d) The check digit is calculated as the last digit of the quantity represented by the expression

$210 - 3(a + c + e + g + i + k) - (b + d + f + h + j)$ where the letters a, b, c, . . . , k represent the 11 successive digits the symbol (e.g., 04200006200 in Figure P6-1) Determine the check digit for the label in Figure P6-1 and check this against the actual bars at the right of that figure.

In April, 1986, Bloomingdale's department store chain announced it would start using the UPC bar code on all items. There is still a question whether department stores will generally adopt the system. Will clothing manufacturers be willing to add the label to their products, regardless of the increase in cost?

(e) Why is there a question about clothing manufacturers when food-product manufacturers comply?

(A monograph describing supermarket automation in much more detail is available from the Department of Technology and Society, State University of New York, Stony Brook, NY 11794. Personal-computer activities have been developed at both Stony Brook and Syracuse University.)

Extra Problem: With a black pen it is easy to change a 5 to an 8 on the left side: we only need to blacken bars five and six. Is it possible to find a change such as the above where the check digit can easily be changed also to be correct?

PROBLEM 7 **SOVIET POSTAL AUTOMATION**

For automatic routing of mail, the Russians have an excellent system. On the bottom left of the envelope, there are guidelines for inserting the "zip code."

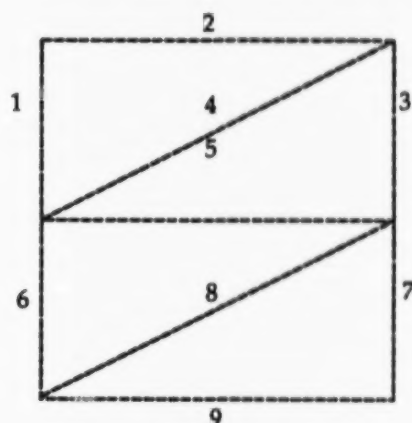


Figure P7-1

For each digit, light guidelines are printed to show nine possible line segments (numbered above). Each decimal digit is represented by the code

	Digit⇒	0	1	2	3	4	5	6	7	8	9
Segment	1	X				X	X			X	X
↓	2	X		X	X		X		X	X	X
	3	X	X	X		X				X	X
	4		X		X			X	X		
	5				X	X	X	X		X	X
	6	X						X	X	X	
	7	X	X			X	X	X		X	
	8			X	X						X
	9	X		X			X	X		X	

Thus, the digit 0 is represented by the nine-digit binary number

111001101

The back of the envelope shows the code to be used — that is, how to form each digit.

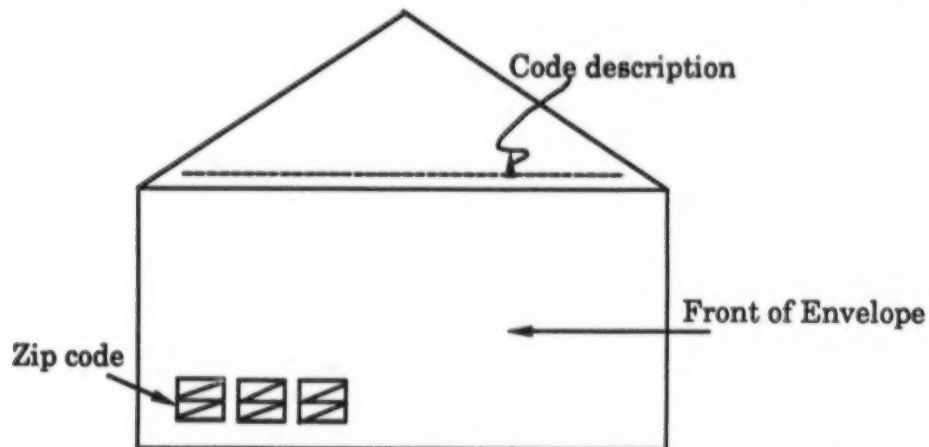


Figure P7-2

The code described above is really not ideal. With nine possible line segments, there are 512 possible binary numbers. We need only ten, one for each decimal digit. We certainly should be able to pick these ten so that each one differs from all others in at least two or three binary digits. But here

0 - 111001101

8 - 111011101

differ in only one digit, the fifth. It is relatively common for the reading equipment to mistake a 0 for an 8, or vice versa. (A careless pencil mark or dirt along segment 5 could cause the error.)

Unfortunately, the Russian engineers felt the above code was the best they could do. Explain why.

Postscript: We show below one scheme where every symbol differs by at least three segments from every other.

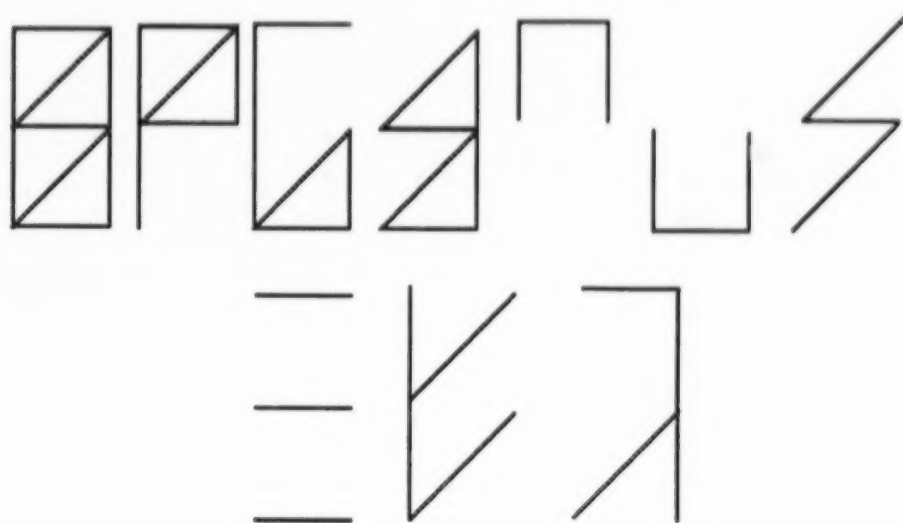


Figure P7-3

PROBLEM 8

ANOTHER CHECK DIGIT

An entirely different scheme leads to the check digit for the ISBN (International Standard Book Number) code. A particular book might have the number

0 - 13 - 925140 - 5

Here

- | | |
|--------|--|
| 0 | The initial zero indicates publication in the U.S. Other nations would have a different initial digit. |
| 13 | These two digits indicate the publisher, in this case Prentice-Hall. |
| 925140 | These six digits are assigned by the publisher to the specific book. |
| 5 | The check digit |

If the ten successive digits are $a, b, c, d, \dots, j$, the check digit j is determined by the fact:

$$10a + 9b + 8c + 7d + 6e + 5f + 4g + 3h + 2i + j$$

must be divisible by 11 with no remainder.

- (a) Show that the check digit for the ISBN number above is actually 5.
- (b) Why would such a complex scheme be used for finding the check digit?

PROBLEM 9

DELTA MODULATION

In Sections 4-6, we used the example of body temperature as a signal to be converted to digital form. We found that there were 25 different temperatures

95 95.5 96 106.5 107

so we needed five bits for each value.

We can get by with fewer bits. Suppose we send the first temperature (98.5) in the usual way (using five bits). Thereafter, we send only the change since the last sample. In our example in Section 5, the temperatures and changes were

97.5
+0.5
98.0
+2
100.0
+2
102.0
0
102.0
-1.5
100.5

Now we might find that there is never a rise of more than 2° or a drop of more than 1.5° . Hence, we need only eight different changes

+2.0
+1.5
+1.0
+0.5

0
-0.5
-1.0
-1.5

These eight values can be represented by only three bits; our digital signal is simpler.

This concept of communicating the change rather than the total value of a signal is called delta modulation (the Greek letter delta is common in mathematics to represent change).

In telephony, delta modulation allows three bits/sample instead of seven. Explain why such a large saving is possible.

ACKNOWLEDGMENT

This material was developed for the Council of Independent Colleges as part of its program in technological literacy under a grant from the Exxon Education Foundation. The author has received significant support from the Alfred P. Sloan Foundation under its New Liberal Arts Program.

John G. Truxal

State University of New York

Stony Brook, NY 11794

For further information about the Council of Independent Colleges
and its project on Technology and the Liberal Arts, write:

CIC
Technology Project
One Dupont Circle, Suite 320
Washington, D.C. 20036

Additional copies of this module can be purchased from
the CIC Publications Department at the address above.